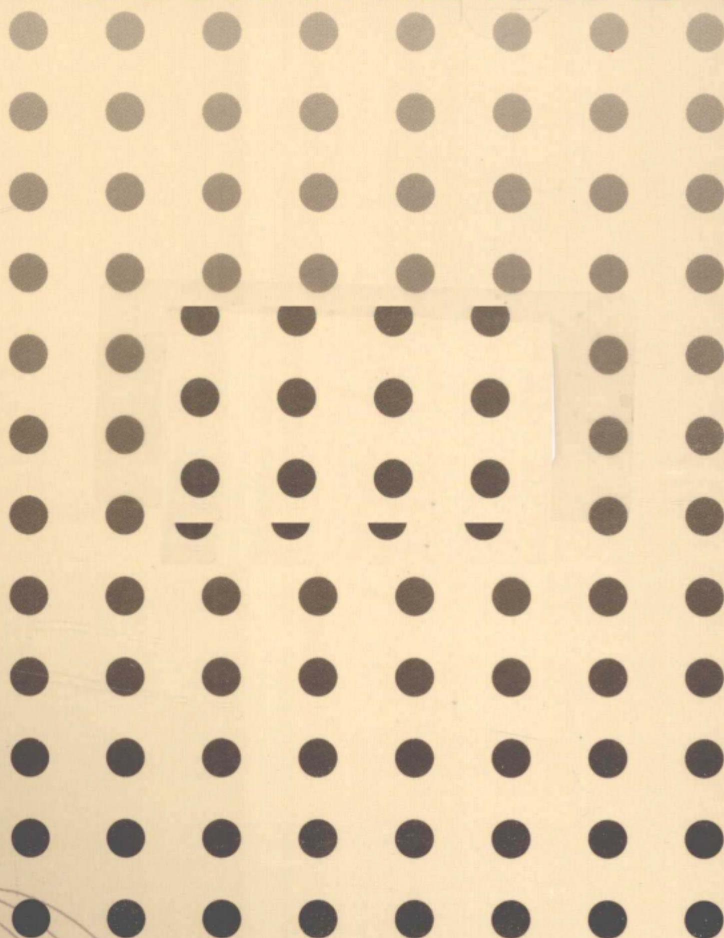



普通高等教育"十一五"规划教材
21世纪研究生数学主干教材
丛书主编　陈　化

曾祥金　张　亮　主编

# 矩阵分析简明教程

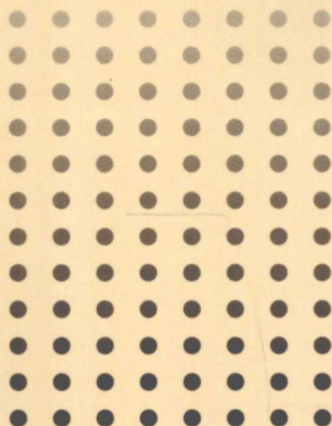
JU ZHEN FENXI JIAN MING JIAO CHENG





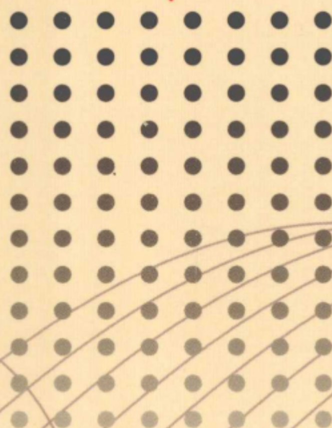
科学出版社
www.sciencep.com

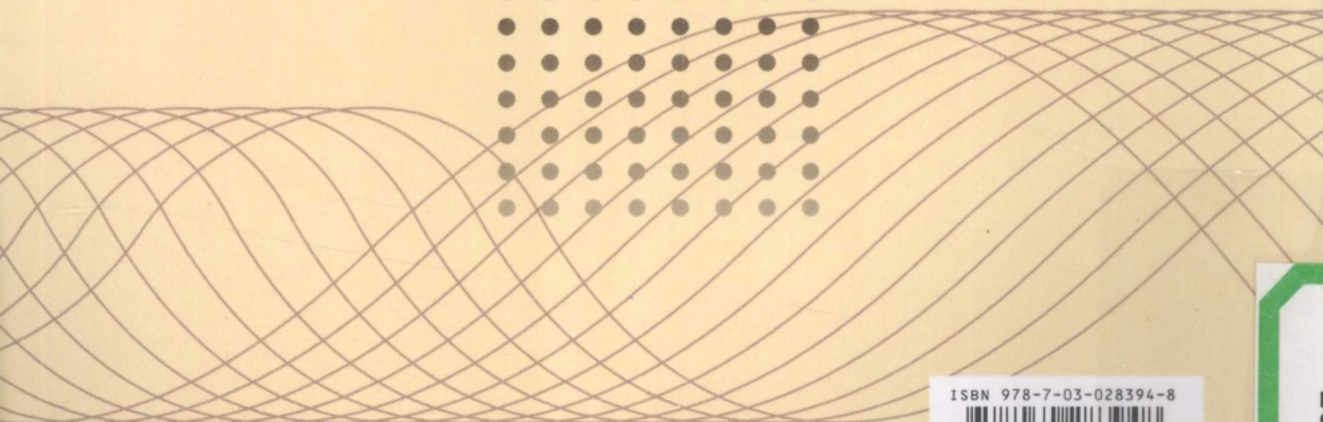
(O-3994.0101)



矩阵分析简明教程










ISBN 978-7-03-028394-8
定价：25.50 元

普通高等教育“十一五”规划教材

21世纪研究生数学主干教材

丛书主编 **陈 化**

# 矩阵分析简明教程

曾祥金 张 亮 主编

科学出版社

北 京

## 内 容 简 介

本书是工科硕士研究生和工程硕士生的教材.全书共分7章，系统地介绍了线性空间和线性变换、内积空间的理论和应用、矩阵的Jordan标准形与若干分解形式、范数理论及其应用、矩阵函数及其应用、特征值的估计与广义逆.各章末配有习题，书末附有答案或提示.本书结合工科的特点，注意理论与应用的结合，引入大量国内外矩阵理论的研究成果，以达到由浅入深，学以致用的目的.

本书也可以供工科高年级本科生、相关教师及工程技术人员阅读或参考.

**图书在版编目(CIP)数据**

矩阵分析简明教程/曾祥金，张亮主编.—北京：科学出版社，2010.8
普通高等教育"十一五"规划教材.21世纪研究生数学主干教材
ISBN 978-7-03-028394-8

Ⅰ.①矩… Ⅱ.①曾…②张… Ⅲ.①矩阵分析-研究生-教材 Ⅳ.①O151.21

中国版本图书馆CIP数据核字(2010)第142474号

责任编辑：吉正霞／责任校对：董艳辉
责任印制：彭 超／封面设计：苏 波

科学出版社 出版
北京东黄城根北街16号
邮政编码：100717
http://www.sciencep.com

武汉市新华印刷有限责任公司印刷

科学出版社发行 各地新华书店经销

*

2010年8月第 一 版 开本：B5(720×1000)
2010年8月第一次印刷 印张：14 1/2
印数：1—3 000 字数：279 000

**定价：25.50元**

(如有印装质量问题，我社负责调换)

# 《21世纪研究生数学主干教材》
# 丛书编委会

# 《21世纪研究生数学主干教材》丛书序

《21世纪研究生数学主干教材》为高等学校研究生数学主干课程系列教材，大致划分为公共数学类、专业数学类两大块，经组编委员会审定，列选科学出版社普通高等教育“十一五”规划教材.

**一、组编机构**

《21世纪研究生数学主干教材》丛书由多所985和211大学联合组编：

**丛书主编** 陈 化

**常务副主编** 刘禄勤

**副 主 编** 吴传生 何 穗 刘安平

**丛书编委**（按姓氏笔画为序）

王卫华 王展青 严国政 杨瑞琰 李 星

肖海军 罗文强 赵东方 黄樟灿 梅全雄

彭 放 彭斯俊 曾祥金 谢民育

**二、编写原则**

**质量**. 质量是图书的生命，保持和发扬科学出版社“三高”、“三严”的传统特色，创造名牌大学和科学出版社双重品牌；适用性是教材的生命力所在，应明确读者对象，篇幅恰当.

**系统**. 研究生教材的系统性高于本科生教材，知识系统，体系完整，逻辑清晰，给学生留下选学和自学的内容和空间.

**创新**. 反映学科发展前沿、先进理念；在知识、内容等方面有所创新，有所贡献，体现教材的知识创新；紧跟和引领教学实践，在教学方法、教材结构、知识组织、详略把握、内容安排上有独到之处，体现教学实践创新.

**三、指导思想**

(1) 力求体系完整，结构严谨，内容精炼，循序渐进，推理简明，深入浅出，富有启发性，让学生打下坚实的理论基础.

(2) 恰当融入现代数学的新思想、新观点、新结果和新方法，使学生有较新的

学术视野.

(3) 为使学生巩固知识和提高应用能力，章末列出习题和思考题，并列出可进一步深入阅读的文献. 书末要给出索引.

(4) 在内容的取舍、叙述的方式、材料的组织安排等方面具有自己的特色.

(5) 公共数学教材注重强化学生的实验训练和实际动手能力，着力培养学生运用现代数学工具(软件)的能力；加强教学内容的应用性，注重案例分析，提高学生对数学知识、数学方法的应用能力及解决问题的能力.

**四、主编职责**

丛书组编委员会和出版社确定全套丛书的编写原则、指导思想和编写规范，在这一框架下，每本教材的主编对本书具有明确的责权利：

1. 拟定指导思想

按照丛书的编写原则和指导思想，拟出编写本书的指导思想和编写说明.

2. 明确特色和编写原则

教材的特色和闪光点；教改、课改动态，学科发展前沿、先进理念如何引入教材；知识和内容创新点及其编写方法；创新与继承的关系把握；教材系统性与教学实践性的关系处理和具体操作；严把教材质量关和适用性.

3. 掌握教材编写环节

(1) 把握教材编写人员水平，原则上要求博士、副教授以上，有多年课程教学经历，熟悉课程和学科领域的发展状况，有教材编写经验，有扎实的文字功底.

(2) 充分注意著作权问题，不侵犯他人著作权.

(3) 讨论、拟定教材提纲，并负责编写组的编写分工、协调与组织.

(4) 拟就内容简介、前言、目录、样章，统稿、定稿，确定交稿时间.

(5) 负责出版事宜，敦促编写组成员使用本教材，并优先选用本系列教材.

《21世纪研究生数学主干教材》组编委员会

2009年10月

# 前　言

矩阵理论自19世纪由凯莱(A. Cayley, 1821—1895)和西尔维斯特(J. J. Sylvester, 1841—1897)创立以来,已经成为一门现代大学的基本数学课程.矩阵理论的出现,对于现代科学技术的发展有着重要的贡献.历史已经证明,矩阵理论是表达量子力学观点的最恰当的语言.科学技术发展到现在,矩阵理论在各个学科领域,如控制理论、优化理论、力学、经济管理、金融等,都有着非常重要的应用.矩阵的方法已经成为现代科技领域不可或缺的研究工具.由于这一原因,我国工科院校已经把"矩阵分析"作为各专业硕士研究生的学位课.本书就是作者在多年为工科硕士研究生讲授该课程的基础上,参考国内外相关教材,结合工科课程的特点,经过多番锤炼编写而成的.

本书内容共分为7章,包括:线性空间与线性变换,内积空间,矩阵的标准形,矩阵分解,范数理论及其应用,矩阵分析及其应用,矩阵特征值的界和非负矩阵.本书前三章内容是线性代数课程的衔接与延伸,这是该课程略为抽象的部分;本书其余各章吸收了近年来国内外最新的具有重大影响的研究成果,譬如Householder分解、Moore-Penrose广义逆、Geršgorin定理、Courant-Fischer定理等.

本书编写力求做到循序渐进,通俗易懂,简洁适用.在理论的论述上本书力求简洁,不苛求对于每一个定理予以证明,但求读者对于理论的精髓能快速的理解.基本方法的介绍力求兼顾应用背景和具体应用,并多用例题加以说明,便于读者迅速掌握.为此,本书根据目前工科硕士研究生的实际需要精心挑选与设计课程内容,每章都配备了一定数量的习题,便于读者进一步加深对课程内容的理解.

本书由曾祥金、张亮主编,吴华安、柳贵平任副主编.在编写的过程中,吴传生教授提出了宝贵的意见,在此表示衷心的感谢.

限于作者水平,书中难免有疏漏之处,敬请读者批评指正.

编　者

2010年5月

# 目 录

序

前言

第 1 章 线性空间与线性变换 …… 001
1.1 线性空间的基本概念 …… 001
1.2 子空间与维数定理 …… 009
1.3 线性空间的同构 …… 014
1.4 线性变换及其矩阵表示 …… 016
习题 1 …… 026

第 2 章 内积空间 …… 029
2.1 内积与欧氏空间 …… 029
2.2 欧氏空间的正交基 …… 033
2.3 欧氏空间的同构 …… 036
2.4 正交补 …… 037
2.5 正交变换 …… 041
2.6 酉空间(复内积空间)简介 …… 044
2.7 正规变换与正规矩阵 …… 046
习题 2 …… 053

第 3 章 矩阵的标准形 …… 055
3.1 Jordan 标准形 …… 055
3.2 λ-矩阵及其 Smith 标准形 …… 063
3.3 Cayley-Hamilton 定理与矩阵的最小多项式 …… 072
习题 3 …… 080

**第 4 章 矩阵分解** …… 083

4.1 矩阵的 LU 分解 …… 083

4.2 矩阵的 QR 分解 …… 089

4.3 矩阵的满秩分解 …… 096

4.4 矩阵的奇异值分解 …… 099

4.5 广义逆矩阵 …… 101

习题 4 …… 106

**第 5 章 范数理论及其应用** …… 108

5.1 向量范数 …… 108

5.2 矩阵范数 …… 116

5.3 范数的应用 …… 120

习题 5 …… 130

**第 6 章 矩阵分析及其应用** …… 132

6.1 矩阵序列与矩阵级数 …… 132

6.2 矩阵函数及其计算 …… 142

6.3 矩阵的微分与积分 …… 151

6.4 矩阵函数的应用 …… 158

习题 6 …… 167

**第 7 章 矩阵特征值的界 非负矩阵** …… 170

7.1 Geršgorin 定理 …… 170

7.2 特征值估计的基本不等式 …… 174

7.3 Courant-Fischer 定理和 Hermite 矩阵的特征值 …… 176

7.4 正矩阵 …… 181

7.5 非负矩阵 …… 184

7.6 随机矩阵 …… 187

7.7 M 矩阵 …… 189

习题 7 …… 194

**习题答案与提示** …… 196

**参考文献** …… 221

# 第1章 线性空间与线性变换

线性空间的概念源自我们所熟悉的向量及其相关运算性质. 将类似的具有共同运算规律的数学对象进行一般的数学描述就得到抽象的线性空间的定义. 因此，线性空间的理论在自然科学、工程技术，特别是数学的其他领域都有着广泛的应用.

## 1.1 线性空间的基本概念

在线性代数中，为研究齐次线性方程组的解的结构，我们学习了实向量空间的基本理论. 在工程技术和科学计算以及数学领域的不同场合，有许多集合本身所伴随的运算具有与实向量空间中的运算相同的本质特征，因而我们有必要将实向量空间的理论进行推广.

我们首先约定，以后所称的数域 $F$，是指实数域或复数域. 在抽象代数中，数域是指至少包含 0 和 1 的数集，在该集中进行的数的和、差、积和商(除数不为 0)的运算是封闭的.

**定义 1.1.1** 设 $V$ 是一个非空集合，其中的元素称为向量，$F$ 是数域，其中的数称为纯量. 在 $V$ 中有向量的加法，使得对任意的向量 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in V$，有和向量 $\boldsymbol{\alpha}+\boldsymbol{\beta} \in V$. 对每个纯量 $x \in F$ 及向量 $\boldsymbol{\alpha} \in V$，有纯量积 $x\boldsymbol{\alpha} \in V$.

如果关于向量加法有运算律：

(i) $\boldsymbol{\alpha}+\boldsymbol{\beta}=\boldsymbol{\beta}+\boldsymbol{\alpha}$;

(ii) $(\boldsymbol{\alpha}+\boldsymbol{\beta})+\boldsymbol{\gamma}=\boldsymbol{\alpha}+(\boldsymbol{\beta}+\boldsymbol{\gamma})$;

(iii) 存在零向量 $\mathbf{0} \in V$，使得对每个 $\boldsymbol{\alpha} \in V$, $\boldsymbol{\alpha}+\mathbf{0}=\boldsymbol{\alpha}$;

(iv) 对每个 $\boldsymbol{\alpha} \in V$，存在负向量 $-\boldsymbol{\alpha}$，使得 $\boldsymbol{\alpha}+(-\boldsymbol{\alpha})=\mathbf{0}$.

关于纯量积有运算律：$x, y \in F$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in V$.

(v) $x(y\boldsymbol{\alpha})=(xy)\boldsymbol{\alpha}$;

(vi) $1\boldsymbol{\alpha}=\boldsymbol{\alpha}$;

(vii) $x(\boldsymbol{\alpha}+\boldsymbol{\beta})=x\boldsymbol{\alpha}+x\boldsymbol{\beta}$;

(viii) $(x+y)\boldsymbol{\alpha}=x\boldsymbol{\alpha}+y\boldsymbol{\alpha}$,

则称 $V$ 是 $F$ 上的一个线性空间(或向量空间).

**例 1.1.1** 在线性代数中,定义的实向量空间 $\mathbf{R}^n$ 关于通常的实向量的加法运算和实数乘向量的数乘运算构成 $\mathbf{R}$ 上的一个线性空间.

**例 1.1.2** 设 $\mathbf{R}[x]_n$ 表示次数小于 $n$ 的多项式的全体(且包含零多项式)对于多项式的加法和数与多项式的乘法构成 $\mathbf{R}$ 上的线性空间.

**例 1.1.3** 设全体 $m\times n$ 复矩阵所成之集为 $\mathbf{C}^{m\times n}$ 按照矩阵的加法和数乘矩阵的运算构成 $\mathbf{C}$ 上的线性空间.

**例 1.1.4** 二阶齐次线性微分方程 $y''+P(x)y'+Q(x)y=0$ 的解全体按通常函数的加法与数乘构成线性空间.

**例 1.1.5** 在正实数集 $\mathbf{R}^+$ 中定义加法$\oplus$和数乘$\circ$如下:

$$a\oplus b=ab$$
$$k\circ a=a^k$$
$$\forall a,\ b\in\mathbf{R}^+,\ k\in\mathbf{R}$$

则 $\mathbf{R}^+$ 是数域 $\mathbf{R}$ 上的线性空间.

线性空间的例子还可以举很多,不过需要指出的是,当一个给定的非空集合在规定的运算下构成线性空间时,可能在另外规定的运算下不能成为线性空间. 例如在 $\mathbf{R}^2$ 中定义

$$(a_1,\ a_2)+(b_1,\ b_2)=(a_1+b_1-1,\ a_2+b_2)$$
$$\lambda(a_1,\ a_2)=(\lambda a_1,\ \lambda a_2)$$

则当 $\lambda\neq 1$ 时,

$$\begin{aligned}\lambda[(a_1,\ a_2)+(b_1,\ b_2)]&=\lambda(a_1+b_1-1,\ a_2+b_2)\\&=(\lambda a_1+\lambda b_1-\lambda,\ \lambda a_2+\lambda b_2)\end{aligned}$$
$$\lambda(a_1,\ a_2)+\lambda(b_1,\ b_2)=(\lambda a_1+\lambda b_1-1,\ \lambda a_2+\lambda b_2)$$

以上两式不相等,说明线性空间定义中的运算公理(vii)不满足. 所以 $\mathbf{R}^2$ 关于以上的运算不是线性空间.

由线性空间的定义可以证明:线性空间 $V$ 中的零向量是唯一的,$V$ 中每个元素的负向量也是唯一的. 并且有以下一些基本性质:

**性质 1.1.1** $0\boldsymbol{\alpha}=\mathbf{0}$, $x\mathbf{0}=\mathbf{0}$.

**性质 1.1.2** 当 $x\boldsymbol{\alpha}=\mathbf{0}$ 时,则 $x=0$ 或 $\boldsymbol{\alpha}=\mathbf{0}$.

**性质 1.1.3** $x(-\boldsymbol{\alpha})=-x\boldsymbol{\alpha}=(-x)\boldsymbol{\alpha}$.

**性质 1.1.4** $(x-y)\boldsymbol{\alpha}=x\boldsymbol{\alpha}-y\boldsymbol{\alpha}$.

**性质 1.1.5** $x(\boldsymbol{\alpha}-\boldsymbol{\beta})=x\boldsymbol{\alpha}-x\boldsymbol{\beta}$.

**证**　我们只证明性质 1.1.3 和性质 1.1.4,其余的留作习题.

根据性质 1.1.1,我们有

$$x\boldsymbol{\alpha}+x(-\boldsymbol{\alpha})=x[\boldsymbol{\alpha}+(-\boldsymbol{\alpha})]=x\mathbf{0}=\mathbf{0}$$

这说明 $x(-\boldsymbol{\alpha})$ 是 $x\boldsymbol{\alpha}$ 的负向量,故 $x(-\boldsymbol{\alpha})=-x\boldsymbol{\alpha}$. 同理可证 $(-x)\boldsymbol{\alpha}=-x\boldsymbol{\alpha}$. 这就证明了性质 1.1.3.

为证性质 1.1.4,首先可以定义向量的减法. 对给定的两个向量 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$,若存在向量 $\boldsymbol{\gamma}$,使得 $\boldsymbol{\alpha}=\boldsymbol{\beta}+\boldsymbol{\gamma}$, 则称向量 $\boldsymbol{\gamma}$ 为向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的差,记为 $\boldsymbol{\gamma}=\boldsymbol{\alpha}-\boldsymbol{\beta}$. 由于 $\boldsymbol{\beta}+[\boldsymbol{\alpha}+(-\boldsymbol{\beta})]=\boldsymbol{\alpha}+[\boldsymbol{\beta}+(-\boldsymbol{\beta})]=\boldsymbol{\alpha}$, 所以有

$$\boldsymbol{\alpha}-\boldsymbol{\beta}=\boldsymbol{\alpha}+(-\boldsymbol{\beta})$$

由性质 1.1.3,在 $(-y)\boldsymbol{\alpha}=-y\boldsymbol{\alpha}$ 两边同加 $x\boldsymbol{\alpha}$,得

$$x\boldsymbol{\alpha}-y\boldsymbol{\alpha}=x\boldsymbol{\alpha}+(-y\boldsymbol{\alpha})=x\boldsymbol{\alpha}+(-y)\boldsymbol{\alpha}=(x-y)\boldsymbol{\alpha}$$

所以性质 1.1.4 成立. □

在线性代数中,我们有关于向量的线性组合、线性相关与线性无关的定义与相关结论,这些定义都可以推广到一般线性空间,而对应的结论也可以证明在线性空间都是成立的.

**定义 1.1.2**　设 $V$ 是数域 $F$ 的线性空间,如果 $V$ 中存在一个有限元素集 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 满足:

(i) $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 线性无关;

(ii) $V$ 中任一向量都可以由 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 线性表示,

则 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 称为 $V$ 的一组基,并称 $V$ 为 $n$ 维线性空间,记为 $n=\dim V$.

当 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 是 $V$ 的基时,对任一 $\boldsymbol{\alpha}\in V$, 必有 $x_1, x_2, \cdots, x_n\in F$, 使得

$$\boldsymbol{\alpha}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n$$

此时,我们称 $n$ 元有序数组 $(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 为向量 $\boldsymbol{\alpha}$ 在基 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 下的坐标.

以下定理表明,向量 $\boldsymbol{\alpha}$ 在给定基下的坐标是唯一确定的.

**定理 1.1.1**　设 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 是线性空间 $V$ 的基,则 $V$ 的任一向量 $\boldsymbol{\alpha}$ 可以由基元素唯一表示.

**证**　若

$$\boldsymbol{\alpha}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n=y_1\boldsymbol{\alpha}_1+y_2\boldsymbol{\alpha}_2+\cdots+y_n\boldsymbol{\alpha}_n$$

则

$$(x_1 - y_1)\boldsymbol{\alpha}_1 + (x_2 - y_2)\boldsymbol{\alpha}_2 + \cdots + (x_n - y_n)\boldsymbol{\alpha}_n = \mathbf{0}$$

由 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 的线性无关性可知 $x_i = y_i$, $i = 1, 2, \cdots, n$. □

线性空间的零子空间的维数规定为 0. 如果一个线性空间中有任意多个线性无关的元素，则称该线性空间为无限维线性空间. 本书只讨论有限维线性空间.

例如，$\mathbf{R}^n$ 是 $n$ 维线性空间，

$$\boldsymbol{e}_1 = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \boldsymbol{e}_2 = \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix}, \cdots, \boldsymbol{e}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix}$$

就是 $\mathbf{R}^n$ 的一组基.

在线性空间 $\mathbf{C}^{m\times n}$ 中，取 $\boldsymbol{E}_{ij}$ 为第 $i$ 行，第 $j$ 列元素为 1，其余元素为 0 的 $m\times n$ 矩阵，$i = 1, 2, \cdots, m$; $j = 1, 2, \cdots, n$，则 $\{\boldsymbol{E}_{ij}\}$ 就是 $\mathbf{C}^{m\times n}$ 的一组基.

容易证明，$n$ 维线性空间 $V$ 中的任意 $n$ 个线性无关的向量所成之集必为 $V$ 的一组基. 任意两组基是等价的向量组.

以下定理告诉我们，如何找到有限维线性空间的基.

**定理 1.1.2** (基的扩张定理)设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r (1 \leqslant r \leqslant n)$ 是 $n$ 维线性空间 $V$ 中的 $r$ 个线性无关的向量，则必有 $V$ 中的 $n-r$ 个向量 $\boldsymbol{\alpha}_{r+1}, \boldsymbol{\alpha}_{r+2}, \cdots, \boldsymbol{\alpha}_n$ 存在，使得 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 成为 $V$ 的基.

**证** 若 $r = n$，则 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 已是 $V$ 的基. 下设 $r < n$，令

$$\begin{aligned} W &= \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r\} \\ &= \{\boldsymbol{\alpha} \in V \mid \boldsymbol{\alpha} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_r\boldsymbol{\alpha}_r,\ x_i \in F,\ 1 \leqslant i \leqslant r\} \end{aligned}$$

故必有 $\boldsymbol{\alpha}_{r+1} \in V$ 且 $\boldsymbol{\alpha}_{r+1} \overline{\in} W$，使得 $\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1}$ 线性无关. 如若不然，即对每个 $\boldsymbol{\beta} \in V - W$，$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r, \boldsymbol{\beta}$ 线性相关，则根据 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r$ 是线性无关可知 $\boldsymbol{\beta}$ 必可由 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r$ 线性表示，此时就必有 $\boldsymbol{\beta} \in W$. 这与 $\boldsymbol{\beta}$ 的取法矛盾.

如果 $r+1 = n$，那么 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_r, \boldsymbol{\alpha}_{r+1} = \boldsymbol{\alpha}_n$ 就是 $V$ 的基，证明到此结束. 如果 $r+1 < n$，我们继续上面的做法，直到找到 $V$ 中的向量 $\boldsymbol{\alpha}_{r+1}, \cdots, \boldsymbol{\alpha}_n$ 使得 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 线性无关. 由于 $V$ 是 $n$ 维线性空间，所以 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 就必为 $V$ 的基. □

现在我们转而讨论向量在不同基下的坐标之间的关系.

设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 及 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n$ 是线性空间 $V$ 的两个基，$\boldsymbol{\alpha} \in V$，

$$\begin{aligned} \boldsymbol{\alpha} &= x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n \\ &= y_1\boldsymbol{\beta}_1 + y_2\boldsymbol{\beta}_2 + \cdots + y_n\boldsymbol{\beta}_n \end{aligned}$$

用矩阵形式写成

$$\boldsymbol{\alpha}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$$

再设

$$\begin{cases}\boldsymbol{\beta}_1=a_{11}\boldsymbol{\alpha}_1+a_{21}\boldsymbol{\alpha}_2+\cdots+a_{n1}\boldsymbol{\alpha}_n\\\boldsymbol{\beta}_2=a_{12}\boldsymbol{\alpha}_1+a_{22}\boldsymbol{\alpha}_2+\cdots+a_{n2}\boldsymbol{\alpha}_n\\\cdots\cdots\\\boldsymbol{\beta}_n=a_{1n}\boldsymbol{\alpha}_1+a_{2n}\boldsymbol{\alpha}_2+\cdots+a_{nn}\boldsymbol{\alpha}_n\end{cases}$$

写成矩阵形式

$$(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{pmatrix}a_{11}&a_{12}&\cdots&a_{1n}\\a_{21}&a_{22}&\cdots&a_{2n}\\\vdots&\vdots&&\vdots\\a_{n1}&a_{n2}&\cdots&a_{nn}\end{pmatrix}$$

$$=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}$$

我们称方阵 $\boldsymbol{A}$ 为由基$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$到基$\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n\}$的过渡矩阵. 于是有

$$\boldsymbol{\alpha}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=(\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_n)\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$$

$$=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$$

由于向量在给定基$\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n\}$下的坐标是唯一的,所以

$$\begin{pmatrix}x_1\\x_2\\\vdots\\x_n\end{pmatrix}=\boldsymbol{A}\begin{pmatrix}y_1\\y_2\\\vdots\\y_n\end{pmatrix}$$

又因过渡矩阵是可逆的,故由上式得到

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \boldsymbol{A}^{-1} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

总结以上的结果,向量在不同基下的坐标之间的关系可以由下述定理给出.

**定理 1.1.3** 设线性空间 $V$ 中的向量 $\boldsymbol{\alpha}$ 在基 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 下的坐标是 $(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$,在基 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n\}$ 下的坐标是 $(y_1, y_2, \cdots, y_n)^{\mathrm{T}}$,从基 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 到基 $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n\}$ 的过渡矩阵是 $\boldsymbol{A}$,则

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

且 $\boldsymbol{A}$ 是可逆的.

**例 1.1.6** 在矩阵空间 $\mathbf{R}^{2\times 2}$ 中

$$\boldsymbol{E}_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{E}_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$$

$$\boldsymbol{E}_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \quad \boldsymbol{E}_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

为 $\mathbf{R}^{2\times 2}$ 的一组基. 对于 $\mathbf{R}^{2\times 2}$ 的元

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}$$

在这个基下的坐标为 $(1, 2, 3, 4)^{\mathrm{T}}$. 可以另取 $\mathbf{R}^{2\times 2}$ 的一组基

$$\boldsymbol{E}_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \quad \boldsymbol{E}_2 = \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}$$

$$\boldsymbol{E}_3 = \begin{pmatrix} 0 & 0 \\ 1 & 1 \end{pmatrix}, \quad \boldsymbol{E}_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

在这个基下 $\boldsymbol{A}$ 的坐标为 $(1, 1, 1, 1)^{\mathrm{T}}$.

**例 1.1.7** 设 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$ 是线性空间 $V$ 的基,向量组

$$\boldsymbol{\beta}_i = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n) \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{pmatrix} = \sum_{j=1}^{n} a_{ji} \boldsymbol{\alpha}_j, \quad i = 1, 2, \cdots, m$$

作矩阵

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m)$$

则 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m$ 线性相关的充要条件是 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m$ 线性相关.

**证** 设 $\sum_{i=1}^{m} \lambda_i \boldsymbol{\beta}_i = \boldsymbol{0}$, $\lambda_i \in F(i = 1, 2, \cdots, n)$,则

$$\sum_{i=1}^{m} \lambda_i \boldsymbol{\beta}_i = \sum_{i=1}^{m} \lambda_i \sum_{j=1}^{n} a_{ji} \boldsymbol{\alpha}_j$$

$$= \sum_{j=1}^{n} \left( \sum_{i=1}^{m} \lambda_i a_{ji} \right) \boldsymbol{\alpha}_j = \boldsymbol{0}$$

因为 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 是线性无关的,上式表明

$$\sum_{i=1}^{m} \lambda_i a_{ji} = 0, \quad j = 1, 2, \cdots, n$$

用矩阵表示就是

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_m \end{pmatrix} = \boldsymbol{0}$$

则

$$\lambda_1 \boldsymbol{\gamma}_1 + \lambda_2 \boldsymbol{\gamma}_2 + \cdots + \lambda_m \boldsymbol{\gamma}_m = \boldsymbol{0}$$

反之由上式又可以推导出

$$\sum_{i=1}^{m} \lambda_i \boldsymbol{\beta}_i = \boldsymbol{0}$$

这说明 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m$ 线性相关的充要条件就是 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m$ 线性相关.

**例 1.1.8** 在 $\mathbf{R}[x]_4$ 中，$\boldsymbol{\alpha}_1 = 1$，$\boldsymbol{\alpha}_2 = x$，$\boldsymbol{\alpha}_3 = x^2$，$\boldsymbol{\alpha}_4 = x^3$ 为一组基. 设

$$\boldsymbol{\beta}_1 = 1+x+x^2+x^3, \quad \boldsymbol{\beta}_2 = 1+x-x^2-x^3$$

$$\boldsymbol{\beta}_3 = 1-x+x^2-x^3, \quad \boldsymbol{\beta}_4 = 1-x-x^2+x^3$$

(1) 证明：$\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，$\boldsymbol{\beta}_3$，$\boldsymbol{\beta}_4$ 为 $\mathbf{R}[x]_4$ 的一组基；

(2) 求由基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\boldsymbol{\alpha}_3$，$\boldsymbol{\alpha}_4$ 到基 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，$\boldsymbol{\beta}_3$，$\boldsymbol{\beta}_4$ 的过渡矩阵；

(3) 求向量 $\boldsymbol{f} = 1+2x+3x^2+4x^3$ 在基 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，$\boldsymbol{\beta}_3$，$\boldsymbol{\beta}_4$ 下的坐标.

**解** (1) 向量 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，$\boldsymbol{\beta}_3$，$\boldsymbol{\beta}_4$ 在基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\boldsymbol{\alpha}_3$，$\boldsymbol{\alpha}_4$ 下的坐标分别为

$$\boldsymbol{X}_1 = \begin{pmatrix}1\\1\\1\\1\end{pmatrix}, \quad \boldsymbol{X}_2 = \begin{pmatrix}1\\1\\-1\\-1\end{pmatrix}, \quad \boldsymbol{X}_3 = \begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}, \quad \boldsymbol{X}_4 = \begin{pmatrix}1\\-1\\-1\\1\end{pmatrix}$$

令

$$\boldsymbol{A} = \begin{pmatrix}1 & 1 & 1 & 1\\1 & 1 & -1 & -1\\1 & -1 & 1 & -1\\1 & -1 & -1 & 1\end{pmatrix}$$

由于 $|\boldsymbol{A}| = -16$，所以 rank $\boldsymbol{A} = 4$，于是 $\boldsymbol{A}$ 的列向量组线性无关. 根据例 1.1.7 的结论知，$\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，$\boldsymbol{\beta}_3$，$\boldsymbol{\beta}_4$ 线性无关，从而是 $\mathbf{R}[x]_4$ 的一组基.

(2) 由(1)知

$$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4) = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4)\boldsymbol{A}$$

则求由基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，$\boldsymbol{\alpha}_3$，$\boldsymbol{\alpha}_4$ 到基 $\boldsymbol{\beta}_1$，$\boldsymbol{\beta}_2$，$\boldsymbol{\beta}_3$，$\boldsymbol{\beta}_4$ 的过渡矩阵为 $\boldsymbol{A}$.

(3)

$$\boldsymbol{A}^{-1} = \frac{1}{4}\begin{pmatrix}1 & 1 & 1 & 1\\1 & 1 & -1 & -1\\1 & -1 & 1 & -1\\1 & -1 & -1 & 1\end{pmatrix}$$

由

$$\boldsymbol{f} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \boldsymbol{\alpha}_4)\begin{pmatrix}1\\2\\3\\4\end{pmatrix} = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4)\boldsymbol{A}^{-1}\begin{pmatrix}1\\2\\3\\4\end{pmatrix}$$

知向量 $f$ 在基 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \boldsymbol{\beta}_4$ 下的坐标为

$$\boldsymbol{Y}=\boldsymbol{A}^{-1}\begin{pmatrix}1\\2\\3\\4\end{pmatrix}=\frac{1}{4}\begin{pmatrix}10\\-4\\-2\\0\end{pmatrix}$$

## 1.2　子空间与维数定理

对于给定的线性空间 $V$ 来说，$V$ 的非空子集中的元素作为 $V$ 的成员自然也继承了 $V$ 中定义的向量加法和纯量积运算，因而也可能成为一个线性空间.

**定义 1.2.1**　如果向量空间 $V$ 的子集 $U$ 在 $V$ 中规定的向量加法和纯量积运算下是一个向量空间，则称 $U$ 是 $V$ 的子空间.

向量空间的任意子集并不是都能成为子空间的. 比如，对于例 1.1.1 中的向量空间 $\mathbf{R}^n$，取子集

$$U=\{(1, a_2, \cdots, a_n) \mid a_2, a_3, \cdots, a_n \in \mathbf{R}\}$$

由于

$$\begin{aligned}&(1, a_1, a_2, \cdots, a_n)+(1, b_1, b_2, \cdots, b_n)\\=&(2, a_1+b_1, a_2+b_2, \cdots, a_n+b_n) \notin U\end{aligned}$$

所以 $U$ 就不是线性空间.

向量空间的子集是否为向量空间，可以用下面的定理判断.

**定理 1.2.1**　设 $V$ 是数域 $F$ 上的向量空间，$U$ 是 $V$ 的一个非空子集，则 $U$ 是 $V$ 的子空间的充分必要条件是

(i) 对任意的 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in U$，有 $\boldsymbol{\alpha}+\boldsymbol{\beta} \in U$；

(ii) 对任意的 $x \in F$，$\boldsymbol{\alpha} \in U$，有 $x\boldsymbol{\alpha} \in U$.

**证**　若 $U$ 是向量空间 $V$ 的子空间，则根据线性空间的定义，在 $U$ 中的向量加法与纯量积运算是封闭的，故条件(i)和条件(ii)是成立的.

反之，当条件(i)和条件(ii)成立时，由 $U$ 中向量的加法运算和纯量积运算的封闭性可知，在 $V$ 中成立的关于这两种运算的八条运算律在 $U$ 中也成立，因而 $U$ 是 $F$ 上的线性空间，从而是 $V$ 的子空间. □

**例 1.2.1**　在几何空间 $\mathbf{R}^3$ 中取过原点的直线

$$L=\{(x, y, z) \mid x=at, y=bt, z=ct, -\infty<t<+\infty\}$$

容易验证 $L$ 是 $\mathbf{R}^3$ 的子空间.

$\mathbf{R}^3$ 中不经过原点的直线不是子空间,因其不含零向量.

**例 1.2.2** 设 $V$ 是一个线性空间, $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s \in V$,则 $W=\{x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_s\boldsymbol{\alpha}_s \mid x_i \in F\}$ 是 $V$ 的子空间. 称为由 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s$ 生成的子空间,记为

$$\mathrm{Span}(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s)$$

**例 1.2.3** 设 $\boldsymbol{A}$ 是 $m\times n$ 实矩阵,则齐次线性方程组 $\boldsymbol{Ax}=\boldsymbol{0}$ 的所有解向量构成 $\mathbf{R}^n$ 的子空间. 我们将这个子空间称为矩阵 $\boldsymbol{A}$ 的核,记为 $\mathrm{Ker}(\boldsymbol{A})$.

令

$$\mathrm{Im}(\boldsymbol{A})=\{\boldsymbol{Ax}\in\mathbf{R}^m \mid \boldsymbol{x}\in\mathbf{R}^n\}$$

则 $\mathrm{Im}(\boldsymbol{A})$是 $\mathbf{R}^m$ 的子空间,称为矩阵 $\boldsymbol{A}$ 的像.

前面我们介绍了线性空间的子空间的概念. 这里将要讨论子空间的两种重要运算——交与和.

**定理 1.2.2** 设 $V_1, V_2$ 是数域 $F$ 上线性空间 $V$ 的两个子空间,则 $V_1\cap V_2$ 也是 $V$ 的子空间.

**证** 因为子空间必含 $V$ 的零向量,所以 $V_1\cap V_2$ 至少有一个零向量,从而 $V_1\cap V_2\neq\varnothing$.

现任取 $\boldsymbol{\alpha}, \boldsymbol{\beta}\in V_1\cap V_2$, $\lambda, \mu\in F$, 则因 $V_1, V_2$ 是子空间,故

$$\lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}\in V_1, \qquad \lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}\in V_2$$

所以

$$\lambda\boldsymbol{\alpha}+\mu\boldsymbol{\beta}\in V_1\cap V_2$$

从而 $V_1\cap V_2$ 是 $V$ 的子空间. □

**例 1.2.4** 在 $\mathbf{R}^3$ 中,取 $V_1, V_2$ 是过原点的平面, $V_1\neq V_2$,则 $V_1\cap V_2$ 就是过原点的一条直线.

**定理 1.2.3** 设 $V_1, V_2$ 是数域 $F$ 上的线性空间 $V$ 的两个子空间,令

$$V_1+V_2=\{\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2 \mid \boldsymbol{\alpha}_1\in V_1, \boldsymbol{\alpha}_2\in V_2\}$$

则 $V_1+V_2$,也是 $V$ 的子空间.

**证** $V_1+V_2$ 显然是非空的. 现在任取 $\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2, \boldsymbol{\beta}_1+\boldsymbol{\beta}_2\in V_1+V_2$, $\lambda, \mu\in F$, 由于 $V_1, V_2$ 是子空间,故 $\lambda\boldsymbol{\alpha}_1+\mu\boldsymbol{\beta}_1\in V_1$, $\lambda\boldsymbol{\alpha}_2+\mu\boldsymbol{\beta}_2\in V_2$. 又

$$\lambda(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)+\mu(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)=(\lambda\boldsymbol{\alpha}_1+\mu\boldsymbol{\beta}_1)+(\lambda\boldsymbol{\alpha}_2+\mu\boldsymbol{\beta}_2)$$

故

$$\lambda(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2)+\mu(\boldsymbol{\beta}_1+\boldsymbol{\beta}_2)\in V_1+V_2$$

所以，$V_1+V_2$ 是 $V$ 的子空间. □

我们称 $V_1+V_2$ 为子空间 $V_1$ 与 $V_2$ 的和.

不难验证，子空间的和运算有下列运算律：

(i)（交换律）$V_1+V_2=V_2+V_1$；

(ii)（结合律）$(V_1+V_2)+V_3=V_1+(V_2+V_3)$.

子空间的交与和运算可以推广到有限个子空间的情形.

**例 1.2.5**　设 $V_1$，$V_2$ 是线性空间 $V$ 的子空间，且

$$V_1=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_s\}$$

$$V_2=\mathrm{Span}\{\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_t\}$$

则

$$V_1+V_2=\mathrm{Span}\{\boldsymbol{\alpha}_1,\cdots,\boldsymbol{\alpha}_s,\boldsymbol{\beta}_1,\cdots,\boldsymbol{\beta}_t\}$$

**定义 1.2.2**　设 $V_1$，$V_2$ 是线性空间 $V$ 的子空间，若 $V_1\cap V_2=\{\mathbf{0}\}$，则 $V_1+V_2$ 称为 $V_1$ 与 $V_2$ 的直和，记为 $V_1\oplus V_2$.

**定理 1.2.4**　$V_1+V_2$ 是直和，当且仅当对每个 $\boldsymbol{\alpha}\in V_1+V_2$，存在唯一的 $\boldsymbol{\alpha}_1\in V_1$，$\boldsymbol{\alpha}_2\in V_2$，使得 $\boldsymbol{\alpha}=\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2$.

**证**　设 $V_1+V_2=V_1\oplus V_2$，若存在 $\boldsymbol{\alpha}\in V_1+V_2$ 有两种表示：

$$\boldsymbol{\alpha}=\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2=\boldsymbol{\alpha}_1'+\boldsymbol{\alpha}_2',\quad \boldsymbol{\alpha}_1,\boldsymbol{\alpha}_1'\in V_1,\ \boldsymbol{\alpha}_2,\boldsymbol{\alpha}_2'\in V_2$$

则

$$\boldsymbol{\alpha}_1-\boldsymbol{\alpha}_1'=\boldsymbol{\alpha}_2'-\boldsymbol{\alpha}_2\in V_1\cap V_2$$

但 $V_1\cap V_2=\{\mathbf{0}\}$，所以

$$\boldsymbol{\alpha}_1=\boldsymbol{\alpha}_1',\qquad \boldsymbol{\alpha}_2=\boldsymbol{\alpha}_2'$$

反之，设每个 $\boldsymbol{\alpha}\in V_1+V_2$ 只有一种表示. 若 $V_1\cap V_2\neq\{\mathbf{0}\}$，取 $\boldsymbol{\alpha}\in V_1\cap V_2$ 且 $\boldsymbol{\alpha}\neq\mathbf{0}$，那么

$$\boldsymbol{\alpha}+(-\boldsymbol{\alpha})=\mathbf{0}+\mathbf{0}=\mathbf{0}$$

上式说明在 $V_1+V_2$ 中有零向量的表示是不唯一的，这与条件矛盾. 故 $V_1\cap V_2=\{\mathbf{0}\}$，即 $V_1+V_2$ 是直和. □

现在研究子空间及由子空间的运算所产生的空间的维数.

**定理 1.2.5**　（维数公式）设 $V_1$ 与 $V_2$ 是数域 $F$ 上线性空间 $V$ 的两个子空间，则

$$\dim(V_1+V_2)=\dim V_1+\dim V_2-\dim(V_1\cap V_2)$$

**证** 设 $\dim V_1=n_1$，$\dim V_2=n_2$，$\dim(V_1\cap V_2)=m$，我们要证明

$$\dim(V_1+V_2)=n_1+n_2-m$$

取 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m$ 为 $V_1\cap V_2$ 的基，由定理 1.1.2，将其依次扩充为 $V_1$，$V_2$ 的基

$$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_{n_1-m}$$

$$\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_{n_2-m}$$

即

$$V_1=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_{n_1-m}\}$$

$$V_2=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_{n_2-m}\}$$

所以

$$V_1+V_2=\mathrm{Span}\{\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\beta}_1,\boldsymbol{\beta}_2,\cdots,\boldsymbol{\beta}_{n_1-m},\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_{n_2-m}\}$$

设

$$\begin{aligned}&k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_m\boldsymbol{\alpha}_m+p_1\boldsymbol{\beta}_1+p_2\boldsymbol{\beta}_2+\cdots+p_{n_1-m}\boldsymbol{\beta}_{n_1-m}\\&\quad+q_1\boldsymbol{\gamma}_1+q_2\boldsymbol{\gamma}_2+\cdots+q_{n_2-m}\boldsymbol{\gamma}_{n_2-m}=\mathbf{0}\end{aligned}$$

令

$$\begin{aligned}\boldsymbol{\alpha}&=k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_m\boldsymbol{\alpha}_m+p_1\boldsymbol{\beta}_1+p_2\boldsymbol{\beta}_2+\cdots+p_{n_1-m}\boldsymbol{\beta}_{n_1-m}\\&=-q_1\boldsymbol{\gamma}_1-q_2\boldsymbol{\gamma}_2-\cdots-q_{n_2-m}\boldsymbol{\gamma}_{n_2-m}\end{aligned}$$

即有 $\boldsymbol{\alpha}\in V_1$ 且 $\boldsymbol{\alpha}\in V_2$，所以 $\boldsymbol{\alpha}\in V_1\cap V_2$，故可令

$$\boldsymbol{\alpha}=l_1\boldsymbol{\alpha}_1+l_2\boldsymbol{\alpha}_2+\cdots+l_m\boldsymbol{\alpha}_m$$

则

$$l_1\boldsymbol{\alpha}_1+l_2\boldsymbol{\alpha}_2+\cdots+l_m\boldsymbol{\alpha}_m=-q_1\boldsymbol{\gamma}_1-q_2\boldsymbol{\gamma}_2-\cdots-q_{n_2-m}\boldsymbol{\gamma}_{n_2-m}$$

即

$$l_1\boldsymbol{\alpha}_1+l_2\boldsymbol{\alpha}_2+\cdots+l_m\boldsymbol{\alpha}_m+q_1\boldsymbol{\gamma}_1+q_2\boldsymbol{\gamma}_2+\cdots+q_{n_2-m}\boldsymbol{\gamma}_{n_2-m}=\mathbf{0}$$

由 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_m,\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\cdots,\boldsymbol{\gamma}_{n_2-m}$ 线性无关知

$$l_1=l_2=\cdots=l_m=q_1=\cdots=q_{n_2-m}=0$$

因而 $\boldsymbol{\alpha}=\mathbf{0}$，从而有

$$k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_m\boldsymbol{\alpha}_m+p_1\boldsymbol{\beta}_1+p_2\boldsymbol{\beta}_2+\cdots+p_{n_1-m}\boldsymbol{\beta}_{n_1-m}=\mathbf{0}$$

由 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_{n_1-m}$线性无关，又得

$$k_1=k_2=\cdots=k_m=p_1=p_2=\cdots=p_{n_1-m}=0$$

这就证明了 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_{n_1-m}, \boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_{n_2-m}$线性无关，因而是 $V_1+V_2$ 的一组基. 于是维数公式成立. □

**推论 1.2.1**　$V_1+V_2$ 是直和的充要条件为

$$\dim(V_1+V_2)=\dim V_1+\dim V_2$$

**推论 1.2.2**　设 $V_1+V_2$ 是直和，若 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{n_1}$ 是 $V_1$ 的基，$\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_{n_2}$ 是 $V_2$ 的基，则 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{n_1}, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_{n_2}$ 是 $V_1+V_2$ 的基.

由此有下述定理.

**定理 1.2.6**　设 $V_1$ 是 $n$ 维线性空间 $V$ 的一个子空间，则一定存在 $V$ 的一个子空间 $V_2$，使 $V=V_1\oplus V_2$.

**证**　设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m$ 是 $V_1$ 的一组基，将其扩充为 $V$ 的基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m, \boldsymbol{\alpha}_{m+1}, \boldsymbol{\alpha}_{m+2}, \cdots, \boldsymbol{\alpha}_n$，令

$$V_2=\mathrm{Span}\{\boldsymbol{\alpha}_{m+1}, \boldsymbol{\alpha}_{m+2}, \cdots, \boldsymbol{\alpha}_n\}$$

显然满足直和维数公式，从而有 $V=V_1\oplus V_2$，即 $V_2$ 为所求. □

定理 1.2.6 表明，有限维线性空间 $V$ 可以作直和分解，且易知这种直和分解不是唯一的.

关于子空间的交与和的概念以及相关定理，可以推广到多个子空间的情形.

**例 1.2.6**　设 $V_1$ 与 $V_2$ 分别是齐次方程组 $x_1+x_2+\cdots+x_n=0$ 与 $x_1=x_2=\cdots=x_n$ 的解空间，证明

$$F^n=V_1\oplus V_2$$

**证**　由于 $x_1+x_2+\cdots+x_n=0$ 的解空间 $V_1$ 是 $n-1$ 维的，其基为

$$\boldsymbol{\alpha}_1=(-1, 1, 0, \cdots, 0)^{\mathrm{T}}$$
$$\boldsymbol{\alpha}_2=(-1, 0, 1, \cdots, 0)^{\mathrm{T}}$$
$$\cdots\cdots$$
$$\boldsymbol{\alpha}_{n-1}=(-1, 0, 0, \cdots, 1)^{\mathrm{T}}$$

而由 $x_1=x_2=\cdots=x_n$ 知 $V_2$ 的维数为 1，其基为

$$\boldsymbol{\beta}=(1, 1, \cdots, 1)^{\mathrm{T}}$$

易知 $\boldsymbol{\beta}, \boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_{n-1}$ 为 $F^n$ 的一组基,从而

$$F^n = V_1 + V_2$$

又

$$\dim F^n = \dim V_1 + \dim V_2$$

故

$$F^n = V_1 \oplus V_2$$

## 1.3 线性空间的同构

从前面的讨论我们已经知道,对 $n$ 维线性空间 $V$,给定一组基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 后,$V$ 中任一向量 $\boldsymbol{\alpha} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n$,且表示法是唯一的. 即

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$(x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in F^n$;反之,对任意$(x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in F^n$,则在 $V$ 中一定存在唯一的一个向量 $x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n = \boldsymbol{\alpha}$,这样,$n$ 维线性空间 $V$ 与 $F^n$之间实质上存在着一个映射,且是双射(既是单射又是满射).

然而,这两个线性空间能否建立某种联系,为我们的研究提供方便?由此,我们给出同构的概念.

**定义 1.3.1** 设 $V_1, V_2$ 都是数域 $F$ 上的线性空间,若存在一个 $V_1$ 到 $V_2$ 的双射 $\sigma$,使得

(i) $\forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in V_1$, $\sigma(\boldsymbol{\alpha}+\boldsymbol{\beta}) = \sigma(\boldsymbol{\alpha}) + \sigma(\boldsymbol{\beta})$;

(ii) $\forall k \in F$, $\boldsymbol{\alpha} \in V$, $\sigma(k\boldsymbol{\alpha}) = k\sigma(\boldsymbol{\alpha})$,

则称 $V_1$ 与 $V_2$ 是同构的,记为 $V_1 \cong V_2$. 这样的映射 $\sigma$ 称为同构映射.

**例 1.3.1** 设 $V$ 是数域 $F$ 上的 $n$ 维线性空间,则 $V \cong F^n$.

**证** 设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 为 $n$ 维线性空间的一组基,那么 $\forall \boldsymbol{\alpha} \in V$ 都有

$$\boldsymbol{\alpha} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n$$

现定义映射 $\sigma$: $V \to F^n$,如下

$$\sigma(\boldsymbol{\alpha}) = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$$

显然 $\sigma$ 是 $V$ 到 $F^n$ 的一个映射.

当 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in V$, $\boldsymbol{\alpha} \neq \boldsymbol{\beta}$ 时,因 $\boldsymbol{\alpha}, \boldsymbol{\beta}$ 在给定基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 下的坐标必定不同,所以 $\sigma$ 是单射.

又 $\forall (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in F^n$, 必有向量

$$x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n = \boldsymbol{\alpha} \in V$$

使

$$\sigma(\boldsymbol{\alpha}) = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$$

从而 $\sigma$ 为满射,故 $\sigma$ 是双射.

$\forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in V$, 则有

$$\boldsymbol{\alpha} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n$$

$$\boldsymbol{\beta} = y_1\boldsymbol{\alpha}_1 + y_2\boldsymbol{\alpha}_2 + \cdots + y_n\boldsymbol{\alpha}_n$$

$$\boldsymbol{\alpha} + \boldsymbol{\beta} = (x_1 + y_1)\boldsymbol{\alpha}_1 + (x_2 + y_2)\boldsymbol{\alpha}_2 + \cdots + (x_n + y_n)\boldsymbol{\alpha}_n$$

故

$$\begin{aligned}\sigma(\boldsymbol{\alpha} + \boldsymbol{\beta}) &= (x_1 + y_1, x_2 + y_2, \cdots, x_n + y_n)^{\mathrm{T}} \\ &= (x_1, x_2, \cdots, x_n)^{\mathrm{T}} + (y_1, y_2, \cdots, y_n)^{\mathrm{T}} \\ &= \sigma(\boldsymbol{\alpha}) + \sigma(\boldsymbol{\beta})\end{aligned}$$

$\forall k \in F$,则有

$$k\boldsymbol{\alpha} = (kx_1)\boldsymbol{\alpha}_1 + (kx_2)\boldsymbol{\alpha}_2 + \cdots + (kx_n)\boldsymbol{\alpha}_n$$

故

$$\begin{aligned}\sigma(k\boldsymbol{\alpha}) &= (kx_1, kx_2, \cdots, kx_n)^{\mathrm{T}} \\ &= k(x_1, x_2, \cdots, x_n)^{\mathrm{T}} \\ &= k\sigma(\boldsymbol{\alpha})\end{aligned}$$

从而 $\sigma$ 是 $V$ 到 $F^n$ 的一个同构映射,即 $V \cong F^n$.

由定义 1.3.1 可以看出,同构映射 $\sigma$ 具有下列性质:

(i) $\sigma(\mathbf{0}) = \mathbf{0}$, $\sigma(-\boldsymbol{\alpha}) = -\sigma(\boldsymbol{\alpha})$;

(ii) $\sigma(k_1\boldsymbol{\alpha}_1 + k_2\boldsymbol{\alpha}_2 + \cdots + k_r\boldsymbol{\alpha}_r) = k_1\sigma(\boldsymbol{\alpha}_1) + k_2\sigma(\boldsymbol{\alpha}_2) + \cdots + k_r\sigma(\boldsymbol{\alpha}_r)$;

(iii) $V$ 中向量组 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s$ 线性相关的充分必要条件是它们的像 $\sigma(\boldsymbol{\alpha}_1)$, $\sigma(\boldsymbol{\alpha}_2)$, $\cdots$, $\sigma(\boldsymbol{\alpha}_s)$ 线性相关.

因为由

$$k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_s\boldsymbol{\alpha}_s=\mathbf{0}$$

可得

$$k_1\sigma(\boldsymbol{\alpha}_1)+k_2\sigma(\boldsymbol{\alpha}_2)+\cdots+k_s\sigma(\boldsymbol{\alpha}_s)=\mathbf{0}$$

反之，由

$$k_1\sigma(\boldsymbol{\alpha}_1)+k_2\sigma(\boldsymbol{\alpha}_2)+\cdots+k_s\sigma(\boldsymbol{\alpha}_s)=\mathbf{0}$$

有

$$\sigma(k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_s\boldsymbol{\alpha}_s)=\mathbf{0}$$

因为 $\sigma$ 是双射，只有 $\sigma(\mathbf{0})=\mathbf{0}$，所以

$$k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_s\boldsymbol{\alpha}_s=\mathbf{0}$$

由此性质可知，同构的线性空间有相同的维数.

同构作为线性空间中的一种关系，满足：

(1) 自反性. 对任何线性空间 $V$，$V\cong V$.

(2) 对称性. 当 $V_1\cong V_2$ 时，必有 $V_2\cong V_1$.

(3) 传递性. 若 $V_1\cong V_2$ 且 $V_2\cong V_3$，则 $V_1\cong V_3$.

由数域 $F$ 上任一 $n$ 维线性空间都与 $F^n$ 同构，及同构性质知，数域 $F$ 上任意两个 $n$ 维线性空间都是同构的.

综上所述，我们有下述定理.

**定理 1.3.1** 数域 $F$ 上两个有限维线性空间同构的充分必要条件是它们有相同的维数.

定理 1.3.1 告诉我们，维数是有限维线性空间的唯一的本质特征.

## 1.4 线性变换及其矩阵表示

**定义 1.4.1** 设 $V$ 是域 $F$ 上的线性空间，映射 $T: V\to V$ 称为线性变换，如果对任意的 $\boldsymbol{\alpha}, \boldsymbol{\beta}\in V$ 和 $\lambda\in F$，有

$$T(\boldsymbol{\alpha}+\boldsymbol{\beta})=T(\boldsymbol{\alpha})+T(\boldsymbol{\beta})$$

$$T(\lambda\boldsymbol{\alpha})=\lambda T(\boldsymbol{\alpha})$$

将 $V$ 中的所有向量映射成零向量的映射必是线性变换，这种线性变换称为零变换，记为 $\mathbf{0}$. 把 $V$ 中每个向量都映射到自身的变换也是线性变换，称为单位变换.

当 $T$ 是 $V$ 的一个线性变换时，记

$$\mathrm{Ker}(T) = \{\boldsymbol{\alpha} \in V \mid T(\boldsymbol{\alpha}) = 0\}$$

$$\mathrm{Im}(T) = \{T(\boldsymbol{\alpha}) \in V \mid \forall \boldsymbol{\alpha} \in V\}$$

易知,$\mathrm{Ker}(T)$和 $\mathrm{Im}(T)$都是 $V$ 的子空间. 这两个子空间分别称为线性变换 $T$ 的核与像.

**例 1.4.1**　设 $V = \mathbf{R}$, $k \in \mathbf{R}$, $k \neq 0$, 定义 $T$: $\mathbf{R} \to \mathbf{R}$ 为

$$T(\boldsymbol{\alpha}) = k\boldsymbol{\alpha}, \quad \boldsymbol{\alpha} \in \mathbf{R}$$

则 $T$ 是 $\mathbf{R}$ 的一个线性变换. 事实上,这一线性变换为一次函数,其图像为一直线.

**例 1.4.2**　在 $\mathbf{R}^2$ 中,作坐标的旋转变换

$$T: \mathbf{R}^2 \to \mathbf{R}^2$$

$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

若记

$$\boldsymbol{\alpha} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \quad \boldsymbol{\beta} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad \mathbf{A} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

则有

$$T(\boldsymbol{\alpha}) = \boldsymbol{\beta} = \boldsymbol{A}\boldsymbol{x}$$

由于

$$\begin{aligned} T(\boldsymbol{\alpha} + \boldsymbol{\alpha}') &= A(\boldsymbol{\alpha} + \boldsymbol{\alpha}') = A\boldsymbol{\alpha} + A\boldsymbol{\alpha}' \\ &= T(\boldsymbol{\alpha}) + T(\boldsymbol{\alpha}'), \quad \boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \mathbf{R}^2 \end{aligned}$$

$$\begin{aligned} T(\lambda\boldsymbol{\alpha}) &= A(\lambda\boldsymbol{\alpha}) = \lambda(A\boldsymbol{\alpha}) \\ &= \lambda T(\boldsymbol{\alpha}), \ \boldsymbol{\alpha} \in \mathbf{R}^2 \end{aligned}$$

故 $T$ 是 $\mathbf{R}^2$ 的一个线性变换.

**例 1.4.3**　令 $V$ 是 $\mathbf{R}$ 上的所有无限次可导实函数的集合,则 $V$ 是线性空间. 定义 $D$: $V \to V$ 为

$$D(f) = f', \quad f \in V$$

那么由导数的运算法则可知 $D$ 是 $V$ 上的线性变换.

线性变换有以下的基本性质.

**定理 1.4.1**　设 $T$: $V \to V$ 是线性变换,则

(i) $T(\mathbf{0}) = \mathbf{0}$;

(ii) $T(-\boldsymbol{\alpha})=-T(\boldsymbol{\alpha})$;

(iii) 对任意的 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_m \in V$ 及 $\lambda_1, \lambda_2, \cdots, \lambda_m \in F$，有

$$T(\lambda_1\boldsymbol{\alpha}_1+\lambda_2\boldsymbol{\alpha}_2+\cdots+\lambda_m\boldsymbol{\alpha}_m)$$
$$=\lambda_1 T(\boldsymbol{\alpha}_1)+\lambda_2 T(\boldsymbol{\alpha}_2)+\cdots+\lambda_m T(\boldsymbol{\alpha}_m)$$

**证** (i) $T(\mathbf{0})=T(0\boldsymbol{\alpha})=0T(\boldsymbol{\alpha})=\mathbf{0}$.

(ii) $T(-\boldsymbol{\alpha})=T[(-1)\boldsymbol{\alpha}]=(-1)T(\boldsymbol{\alpha})=-T(\boldsymbol{\alpha})$.

(iii) $T(\lambda_1\boldsymbol{\alpha}_1+\cdots+\lambda_m\boldsymbol{\alpha}_m)=T(\lambda_1\boldsymbol{\alpha}_1)+T(\lambda_2\boldsymbol{\alpha}_2+\cdots+\lambda_m\boldsymbol{\alpha}_m)=\cdots$ $=T(\lambda_1\boldsymbol{\alpha}_1)+T(\lambda_2\boldsymbol{\alpha}_2)+\cdots+T(\lambda_m\boldsymbol{\alpha}_m)=\lambda_1 T(\boldsymbol{\alpha}_1)+\lambda_2 T(\boldsymbol{\alpha}_2)+\cdots+\lambda_m T(\boldsymbol{\alpha}_m)$.

□

**例 1.4.4** 设 $\boldsymbol{A}\in F^{n\times n}$，且

$$\boldsymbol{A}=\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}=(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)$$

定义映射 $T: F^n \to F^n$ 为

$$T(\boldsymbol{\alpha})=\boldsymbol{A\alpha}, \quad \boldsymbol{\alpha}\in F^n$$

则

(1) $T$ 是线性变换；

(2) $\mathrm{Ker}(T)$ 是 $\boldsymbol{A\alpha}=\mathbf{0}$ 的解空间；

(3) $\mathrm{Im}(T)=\mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$；

(4) $\dim \mathrm{Ker}(T)+\dim \mathrm{Im}(T)=n$.

**证** (1) 是显然的.

(2) $\boldsymbol{\alpha}\in \mathrm{Ker}(T)$ 当且仅当 $T(\boldsymbol{\alpha})=\boldsymbol{A\alpha}=\mathbf{0}$.

(3) 由 $\mathrm{Im}(T)$ 的定义可知，$\boldsymbol{\beta}\in \mathrm{Im}(T)$ 当且仅当存在 $\boldsymbol{\alpha}\in F^n$，使得 $T(\boldsymbol{\alpha})=\boldsymbol{A\alpha}=\boldsymbol{\beta}$. 若 $\boldsymbol{\alpha}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$，则

$$\boldsymbol{A\alpha}=(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=x_1\boldsymbol{\alpha}_1+x_2\boldsymbol{\alpha}_2+\cdots+x_n\boldsymbol{\alpha}_n=\boldsymbol{\beta}$$

即

$$\boldsymbol{\beta}\in \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$$

（4）由线性代数的知识得

$$\dim \mathrm{Ker}(T) = n - R(\boldsymbol{A}) = n - \dim \mathrm{Im}(T)$$

设 $T_1$，$T_2$，$T$ 都是 $V$ 上的线性变换，称 $T_1$ 与 $T_2$ 相等，如果对 $\forall \alpha \in V$，都有 $T_1(\boldsymbol{\alpha}) = T_2(\boldsymbol{\alpha})$. 定义线性变换的和与数乘运算如下：

$$(T_1 + T_2)(\boldsymbol{\alpha}) = T_1(\boldsymbol{\alpha}) + T_2(\boldsymbol{\alpha}), \quad \boldsymbol{\alpha} \in V$$

$$(\lambda T)(\boldsymbol{\alpha}) = \lambda T(\boldsymbol{\alpha}), \quad \lambda \in F$$

容易证明，$T_1 + T_2$，$\lambda T$ 也是线性变换.

令 $L(V)$ 为所有 $V$ 上的线性变换之集，以下结论是显然的.

**定理 1.4.2**　$L(V)$ 是 $F$ 上的线性空间.

从上面得知，线性空间 $V$ 上的所有线性变换组成的集合 $L(V)$，对于线性变换的加法及数量乘法，也构成一个线性空间. 若 $\dim V = n$，那么 $L(V)$ 的维数是多少？$L(V)$ 与我们熟知的哪一个线性空间有关系？

设 $V$ 是数域 $F$ 上的 $n$ 维线性空间，$\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 是 $V$ 的一组基，则空间中任一向量 $\boldsymbol{\alpha}$ 可以唯一地由基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 线性表示，即有关系式

$$\boldsymbol{\alpha} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n$$

设 $T \in L(V)$，则有

$$\begin{aligned} T(\boldsymbol{\alpha}) &= T(x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n) \\ &= x_1 T(\boldsymbol{\alpha}_1) + x_2 T(\boldsymbol{\alpha}_2) + \cdots + x_n T(\boldsymbol{\alpha}_n) \end{aligned}$$

上式表明，如果我们知道了基 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 的像，则空间中任一向量 $\boldsymbol{\alpha}$ 的像就知道了. 即有下述定理.

**定理 1.4.3**　设 $\boldsymbol{\alpha}_1$，$\boldsymbol{\alpha}_2$，…，$\boldsymbol{\alpha}_n$ 是 $n$ 维线性空间 $V$ 的一组基，$T_1$，$T_2 \in L(V)$，则 $T_1 = T_2$ 的充分必要条件是

$$T_1(\boldsymbol{\alpha}_i) = T_2(\boldsymbol{\alpha}_i) \ (i = 1,\ 2,\ \cdots,\ n)$$

**证**　由线性变换相等知，当 $T_1 = T_2$ 时，必有 $T_1(\boldsymbol{\alpha}_i) = T_2(\boldsymbol{\alpha}_i)$ $(i = 1,\ 2,\ \cdots,\ n)$. 现在只要证明对任一向量 $\boldsymbol{\alpha} \in V$，等式 $T_1(\boldsymbol{\alpha}) = T_2(\boldsymbol{\alpha})$ 成立. 由

$$\begin{aligned} T_1(\boldsymbol{\alpha}) &= T_1(x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n) \\ &= x_1 T_1(\boldsymbol{\alpha}_1) + x_2 T_1(\boldsymbol{\alpha}_2) + \cdots + x_n T_1(\boldsymbol{\alpha}_n) \\ &= x_1 T_2(\boldsymbol{\alpha}_1) + x_2 T_2(\boldsymbol{\alpha}_2) + \cdots + x_n T_2(\boldsymbol{\alpha}_n) \\ &= T_2(x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n) \\ &= T_2(\boldsymbol{\alpha}) \end{aligned}$$

结论成立. □

定理 1.4.3 的意义就是,一个线性变换完全被该变换在一组基上的作用所决定. 下面我们进一步说明,基向量的像却完全可以是任意的.

**定理 1.4.4** 设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 是 $n$ 维线性空间 $V$ 的一组基. 对于任意一组向量 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n \in V$, 一定存在 $V$ 上唯一的线性变换 $T$,使

$$T(\boldsymbol{\alpha}_i) = \boldsymbol{\beta}_i, \quad i = 1, 2, \cdots, n$$

**证** 先证存在性. 设 $\forall \boldsymbol{\alpha} \in V$, 则

$$\boldsymbol{\alpha} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n$$

我们定义 $V$ 的变换 $T$ 为

$$T(\boldsymbol{\alpha}) = x_1\boldsymbol{\beta}_1 + x_2\boldsymbol{\beta}_2 + \cdots + x_n\boldsymbol{\beta}_n$$

下面证明 $T$ 是线性变换.

在 $V$ 中任取两个向量 $\boldsymbol{\alpha}, \boldsymbol{\beta}$:

$$\boldsymbol{\alpha} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n$$

$$\boldsymbol{\beta} = y_1\boldsymbol{\alpha}_1 + y_2\boldsymbol{\alpha}_2 + \cdots + y_n\boldsymbol{\alpha}_n$$

于是

$$\boldsymbol{\alpha} + \boldsymbol{\beta} = (x_1 + y_1)\boldsymbol{\alpha}_1 + (x_2 + y_2)\boldsymbol{\alpha}_2 + \cdots + (x_n + y_n)\boldsymbol{\alpha}_n$$

$$k\boldsymbol{\alpha} = (kx_1)\boldsymbol{\alpha}_1 + (kx_2)\boldsymbol{\alpha}_2 + \cdots + (kx_n)\boldsymbol{\alpha}_n$$

故

$$\begin{aligned} T(\boldsymbol{\alpha} + \boldsymbol{\beta}) &= (x_1 + y_1)\boldsymbol{\beta}_1 + (x_2 + y_2)\boldsymbol{\beta}_2 + \cdots + (x_n + y_n)\boldsymbol{\beta}_n \\ &= (x_1\boldsymbol{\beta}_1 + x_2\boldsymbol{\beta}_2 + \cdots + x_n\boldsymbol{\beta}_n) + (y_1\boldsymbol{\beta}_1 + y_2\boldsymbol{\beta}_2 + \cdots + y_n\boldsymbol{\beta}_n) \\ &= T(\boldsymbol{\alpha}) + T(\boldsymbol{\beta}) \end{aligned}$$

$$\begin{aligned} T(k\boldsymbol{\alpha}) &= (kx_1)\boldsymbol{\beta}_1 + (kx_2)\boldsymbol{\beta}_2 + \cdots + (kx_n)\boldsymbol{\beta}_n \\ &= k(x_1\boldsymbol{\beta}_1 + x_2\boldsymbol{\beta}_2 + \cdots + x_n\boldsymbol{\beta}_n) \\ &= kT(\boldsymbol{\alpha}) \end{aligned}$$

因此, $T \in L(V)$. 又由

$$\boldsymbol{\alpha}_i = 0\boldsymbol{\alpha}_1 + \cdots + 0\boldsymbol{\alpha}_{i-1} + 1\boldsymbol{\alpha}_i + 0\boldsymbol{\alpha}_{i+1} + \cdots + 0\boldsymbol{\alpha}_n, \quad i = 1, 2, \cdots, n$$

所以

$$T(\boldsymbol{\alpha}_i) = 0\boldsymbol{\beta}_1 + \cdots + 0\boldsymbol{\beta}_{i-1} + 1\boldsymbol{\beta}_i + 0\boldsymbol{\beta}_{i+1} + \cdots + 0\boldsymbol{\beta}_n = \boldsymbol{\beta}_i$$

唯一性由定理 1.4.3 可得. □

有了上面的准备,现在我们可以研究线性变换和矩阵的关系.

**定义 1.4.2** 设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 是数域 $F$ 上 $n$ 维线性空间 $V$ 的一组基. $T$ 是 $V$ 上的一个线性变换. 基向量的像可以由基线性表示

$$T(\boldsymbol{\alpha}_1) = a_{11}\boldsymbol{\alpha}_1 + a_{21}\boldsymbol{\alpha}_2 + \cdots + a_{n1}\boldsymbol{\alpha}_n$$

$$T(\boldsymbol{\alpha}_2) = a_{12}\boldsymbol{\alpha}_1 + a_{22}\boldsymbol{\alpha}_2 + \cdots + a_{n2}\boldsymbol{\alpha}_n$$

$$\cdots\cdots$$

$$T(\boldsymbol{\alpha}_n) = a_{1n}\boldsymbol{\alpha}_1 + a_{2n}\boldsymbol{\alpha}_2 + \cdots + a_{nn}\boldsymbol{\alpha}_n$$

用矩阵来表示就是

$$\begin{aligned} T(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n) &= (T(\boldsymbol{\alpha}_1), T(\boldsymbol{\alpha}_2), \cdots, T(\boldsymbol{\alpha}_n)) \\ &= (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{A} \end{aligned}$$

其中

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

矩阵 $\boldsymbol{A}$ 称为线性变换 $T$ 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 下的矩阵.

**例 1.4.5** 已知 $\boldsymbol{B} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$,线性空间

$$V = \left\{ \boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix} \middle| \ x_{11} + x_{22} = 0, \ x_{ij} \in \boldsymbol{R} \right\}$$

定义线性变换

$$T(\boldsymbol{X}) = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{X} - \boldsymbol{X}^{\mathrm{T}}\boldsymbol{B}$$

试求 $V$ 的一组基,并求 $T$ 在所求基下的矩阵.

**解** 先求 $V$ 的一组基

$$\boldsymbol{E}_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \boldsymbol{E}_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad \boldsymbol{E}_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$$

则线性变换 $T$ 在基 $\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3$ 下像

$$T(\boldsymbol{E}_1)=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{E}_1-\boldsymbol{E}_1^{\mathrm{T}}\boldsymbol{B}=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}=-\boldsymbol{E}_2+\boldsymbol{E}_3$$

$$T(\boldsymbol{E}_2)=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{E}_2-\boldsymbol{E}_2^{\mathrm{T}}\boldsymbol{B}=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}=\boldsymbol{E}_2-\boldsymbol{E}_3$$

$$T(\boldsymbol{E}_3)=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{E}_3-\boldsymbol{E}_3^{\mathrm{T}}\boldsymbol{B}=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}=-\boldsymbol{E}_2+\boldsymbol{E}_3$$

从而线性变换 $T$ 在基 $\boldsymbol{E}_1$, $\boldsymbol{E}_2$, $\boldsymbol{E}_3$ 下的矩阵

$$\boldsymbol{A}=\begin{bmatrix}0 & 0 & 0\\ -1 & 1 & -1\\ 1 & -1 & 1\end{bmatrix}$$

从定义 1.4.2 知，在取定一组基后，我们建立了由数域 $F$ 上 $n$ 维线性空间上的线性变换 $L(V)$ 到数域 $F$ 上的 $n\times n$ 矩阵的一个映射. 前述定理 1.4.3 说明这个映射是单射，定理 1.4.4 说明这个映射是满射. 即说明在二者之间建立了一个双射. 从而有下述定理.

**定理 1.4.5** 数域 $F$ 上 $n$ 维线性空间 $V$ 的所有线性变换构成的线性空间 $L(V)$，在取定 $V$ 的一组基后，$L(V)$ 与数域 $F$ 上一切 $n\times n$ 矩阵所构成的线性空间 $F^{n\times n}$ 是同构的.

**证** 由前面分析知，只需证明保持线性运算. 设 $\boldsymbol{\alpha}_1$, $\boldsymbol{\alpha}_2$, $\cdots$, $\boldsymbol{\alpha}_n$ 为 $V$ 的一组基，$T_1$, $T_2\in L(V)$，且

$$T_1(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}$$

$$T_2(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{B}$$

则

$$\begin{aligned}(T_1+T_2)(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)&=T_1(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)+T_2(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\\&=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}+(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{B}\\&=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)(\boldsymbol{A}+\boldsymbol{B})\end{aligned}$$

$$\begin{aligned}(kT_1)(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)&=k[T_1(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)]\\&=k(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)\boldsymbol{A}\\&=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n)(k\boldsymbol{A})\end{aligned}$$

故

$$L(V)\cong F^{n\times n}$$

□

**推论 1.4.1** $\dim L(V)=\dim F^{n\times n}=n^2$.

利用线性变换的矩阵可以直接计算一个向量的像.

**定理 1.4.6**　设线性变换 $T$ 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{A}$，向量 $\boldsymbol{\alpha}$ 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 下的坐标为 $(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$，则 $T(\boldsymbol{\alpha})$ 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 下的坐标 $(y_1, y_2, \cdots, y_n)^{\mathrm{T}}$ 为

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

**证**　由条件知

$$\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$T(\boldsymbol{\alpha}) = (T(\boldsymbol{\alpha}_1), T(\boldsymbol{\alpha}_2), \cdots, T(\boldsymbol{\alpha}_n)) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

又

$$T(\boldsymbol{\alpha}) = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n) \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$

从而有

$$\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} = \boldsymbol{A} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

□

最后，研究同一线性变换在不同基下矩阵之间的关系.

**定理 1.4.7** 设 $T$ 是 $n$ 维线性空间 $V$ 上的线性变换，$T$ 在基 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{A}$，在基 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n$ 下的矩阵为 $\boldsymbol{B}$，则矩阵 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似，亦即存在可逆阵 $\boldsymbol{P}$，使得 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{AP}$；反之，如果两个矩阵相似，那么它们可以分别看做同一线性变换在两组基下的矩阵.

**证** 设 $n$ 维线性空间中的线性变换 $T$ 在两组不同基

$$\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$$

$$\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n$$

下的矩阵分别为 $\boldsymbol{A}$ 和 $\boldsymbol{B}$，即

$$T(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)=(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{A}$$

$$T(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n)=(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n)\boldsymbol{B}$$

又设

$$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{P}$$

则有

$$\begin{aligned}T(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n) &= T[(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{P}] \\ &= [T(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)]\boldsymbol{P} \\ &= (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{AP} \\ &= (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n)\boldsymbol{P}^{-1}\boldsymbol{AP}\end{aligned}$$

因此得

$$\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{AP}$$

反之，若 $\boldsymbol{B}=\boldsymbol{P}^{-1}\boldsymbol{AP}$. 设 $T$ 在 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 下的矩阵为 $\boldsymbol{A}$，令

$$(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n)=(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{P}$$

显然，$\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n$ 为基，这时

$$\begin{aligned}T(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n) &= T[(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{P}] \\ &= T(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{P} \\ &= (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\boldsymbol{AP} \\ &= (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n)\boldsymbol{P}^{-1}\boldsymbol{AP} \\ &= (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n)\boldsymbol{B}\end{aligned}$$

□

相似矩阵的概念及一些性质，读者已熟知，在此不赘述.

**例 1.4.6**　对于例 1.4.5 中定义的线性变换 $T$,求线性空间 $V$ 一组基,使得 $T$ 在所求基下的矩阵为对角矩阵.

**解**　在例 1.4.5 中,已经求出线性变换 $T$ 在基 $\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3$ 下的矩阵为

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0 & 0 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

下面将矩阵 $\boldsymbol{A}$ 对角化. 先求 $\boldsymbol{A}$ 的特征值和特征向量.

由 $|\lambda \boldsymbol{E} - \boldsymbol{A}| = \lambda^2(\lambda - 2)$ 得特征值

$$\lambda_1 = \lambda_2 = 0, \qquad \lambda_3 = 2$$

当 $\lambda_1 = \lambda_2 = 0$ 时,解方程组 $(\lambda_1 \boldsymbol{E} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ 的特征向量

$$\boldsymbol{p}_1 = (1, 1, 0)^{\mathrm{T}}, \qquad \boldsymbol{p}_2 = (0, 1, 1)^{\mathrm{T}}$$

当 $\lambda_3 = 2$ 时,求得特征向量

$$\boldsymbol{p}_3 = (0, 1, -1)^{\mathrm{T}}$$

令

$$\boldsymbol{P} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3) = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix}, \qquad \boldsymbol{\Lambda} = \begin{pmatrix} 0 & & \\ & 0 & \\ & & 2 \end{pmatrix}$$

从而有

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{\Lambda}$$

令

$$\begin{aligned} (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3) &= (\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3)\boldsymbol{P} \\ &= (\boldsymbol{E}_1, \boldsymbol{E}_2, \boldsymbol{E}_3)\begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & -1 \end{pmatrix} \\ &= (\boldsymbol{E}_1 + \boldsymbol{E}_2, \boldsymbol{E}_2 + \boldsymbol{E}_3, \boldsymbol{E}_2 - \boldsymbol{E}_3) \end{aligned}$$

即

$$\boldsymbol{\beta}_1 = \boldsymbol{E}_1 + \boldsymbol{E}_2 = \begin{pmatrix} 1 & 1 \\ 0 & -1 \end{pmatrix}$$

$$\boldsymbol{\beta}_2 = \boldsymbol{E}_2 + \boldsymbol{E}_3 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

$$\boldsymbol{\beta}_3 = \boldsymbol{E}_2 - \boldsymbol{E}_3 = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

故 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \boldsymbol{\beta}_3$ 是 $V$ 的一组基，且 $T$ 这个基下的矩阵为对角矩阵 $\boldsymbol{\Lambda}$.



## 习题 1

1. 按通常矩阵的加法及数与矩阵的乘法，下列数域 $F$ 上方阵集合是否构成 $F$ 上的线性空间：

(1) 全体形如 $\begin{pmatrix} 0 & a \\ -a & b \end{pmatrix}$ 的二阶方阵的集合；

(2) 全体 $n$ 阶对称(或反对称、上三角)矩阵的集合；

(3) $V = \{\boldsymbol{X} \mid \boldsymbol{AX} = \boldsymbol{0},\ \boldsymbol{X} \in F^{n\times n}\}$ ($\boldsymbol{A}$ 为给定的 $n$ 阶方阵).

2. 在 $n$ 维线性空间 $F^n$ 中，下列 $n$ 维向量的集合 $W$，是否构成 $F^n$ 的子空间：

(1) $W = \left\{(a_1, a_2, \cdots, a_n) \,\middle|\, \sum_{i=1}^{n} a_i = 1\right\}$；

(2) $W = \left\{(a_1, a_2, \cdots, a_n) \,\middle|\, \sum_{i=1}^{n} a_i = 0\right\}$；

(3) $W = \{\boldsymbol{\alpha} = (a_1, a_2, \cdots, a_n)^{\mathrm{T}} \mid \boldsymbol{A\alpha} = \boldsymbol{0},\ \boldsymbol{A} \in F^{n\times n}\}$.

3. 在 $m\times n$ 维线性空间 $\mathbf{R}^{m\times n}$ 中，下列子集 $W$ 是否构成 $\mathbf{R}^{m\times n}$ 的子空间：

(1) $W = \left\{\boldsymbol{A} \,\middle|\, \boldsymbol{A} = (a_{ij})_{mn},\ \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} = 1\right\}$；

(2) $W = \left\{\boldsymbol{A} \,\middle|\, \boldsymbol{A} = (a_{ij})_{mn},\ \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij} = 0\right\}$.

4. 设

$$W = \left\{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in \mathbf{R}^{2\times 2} \,\middle|\, a_{11} + a_{22} = 0\right\}$$

(1) 证明 $W$ 是 $\mathbf{R}^{2\times 2}$ 的子空间；

(2) 试求 $W$ 的一组基；

(3) 试求 $\boldsymbol{A} = \begin{pmatrix} 3 & 2 \\ 5 & -3 \end{pmatrix}$ 在所求基下的坐标.

5. 设 $\boldsymbol{A} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \in \mathbf{R}^{2\times 2}$.

(1) 证明 $C(\boldsymbol{A}) = \{\boldsymbol{B} \in \mathbf{R}^{2\times 2} \mid \boldsymbol{AB} = \boldsymbol{BA}\}$ 是 $\mathbf{R}^{2\times 2}$ 的子空间；

(2) 求 $C(\boldsymbol{A})$ 的维数与一组基.

6. 在 $\mathbf{R}^{2\times2}$ 中, $\boldsymbol{\varepsilon}_1=\begin{pmatrix}1&0\\0&0\end{pmatrix}$, $\boldsymbol{\varepsilon}_2=\begin{pmatrix}0&1\\0&0\end{pmatrix}$, $\boldsymbol{\varepsilon}_3=\begin{pmatrix}0&0\\1&0\end{pmatrix}$, $\boldsymbol{\varepsilon}_4=\begin{pmatrix}0&0\\0&1\end{pmatrix}$ 与 $\boldsymbol{\eta}_1=\begin{pmatrix}-1&0\\0&2\end{pmatrix}$, $\boldsymbol{\eta}_2=\begin{pmatrix}0&3\\-1&4\end{pmatrix}$, $\boldsymbol{\eta}_3=\begin{pmatrix}2&1\\0&1\end{pmatrix}$, $\boldsymbol{\eta}_4=\begin{pmatrix}1&-3\\0&2\end{pmatrix}$ 是两组基.

(1) 试求 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4$ 到 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4$ 的过渡矩阵;

(2) 试求 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4$ 到 $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4$ 的过渡矩阵;

(3) 试求 $\boldsymbol{A}=\begin{pmatrix}-1&3\\0&2\end{pmatrix}$ 在这两组基下的坐标.

7. 试证：在 $\mathbf{R}^4$ 中,由(1, 1, 0, 0), (1, 0, 1, 1)生成的子空间与由(2, −1, 3, 3), (0, 1, −1, −1)生成的子空间相同.

8. 试求 $\mathbf{R}^4$ 的子空间

$$W_1=\{(a_1, a_2, a_3, a_4)\mid a_1-a_2+a_3-a_4=0\}$$

$$W_2=\{(a_1, a_2, a_3, a_4)\mid a_1+a_2+a_3+a_4=0\}$$

的交 $W_1\cap W_2$ 的一组基.

9. 证明：

$$T_1(x_1, x_2)=(x_2, -x_1)$$

$$T_2(x_1, x_2)=(x_1, -x_2)$$

$\forall(x_1, x_2)\in\mathbf{R}^2$, 是 $\mathbf{R}^2$ 上的两个线性变换,并求 $T_1+T_2$.

10. 设 $\boldsymbol{B}=\begin{pmatrix}1&1\\1&-1\end{pmatrix}\in F^{2\times2}$, $\forall\boldsymbol{X}\in F^{2\times2}$ 定义 $T(\boldsymbol{X})=\boldsymbol{XB}$.

(1) 证明 $T$ 是 $F^{2\times2}$ 上的线性变换.

(2) 试求 $T$ 在基 $\boldsymbol{E}_{11}=\begin{pmatrix}1&0\\0&0\end{pmatrix}$, $\boldsymbol{E}_{12}=\begin{pmatrix}0&1\\0&0\end{pmatrix}$, $\boldsymbol{E}_{21}=\begin{pmatrix}0&0\\1&0\end{pmatrix}$, $\boldsymbol{E}_{22}=\begin{pmatrix}0&0\\0&1\end{pmatrix}$ 下的矩阵.

11. 设 $\boldsymbol{B}=\begin{pmatrix}0&1\\4&0\end{pmatrix}\in\mathbf{R}^{2\times2}$, 在 $\mathbf{R}^{2\times2}$ 中定义线性变换

$$T(\boldsymbol{X})=\boldsymbol{XB}$$

试求 $\mathbf{R}^{2\times2}$ 的一个基,使 $T$ 在所求基 $F$ 的矩阵为对角矩阵.

12. 在 $\mathbf{R}^3$ 中,已知线性变换 $T$ 在基

$$\boldsymbol{\eta}_1=(-1, 1, 1)^{\mathrm{T}},\quad \boldsymbol{\eta}_2=(1, 0, -1)^{\mathrm{T}},\quad \boldsymbol{\eta}_3=(0, 0, 1)^{\mathrm{T}}$$

下的矩阵为

$$\begin{pmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 2 & 1 \end{pmatrix}$$

试求 $T$ 在基 $\boldsymbol{\varepsilon}_1 = (1, 0, 0)^{\mathrm{T}}$，$\boldsymbol{\varepsilon}_2 = (0, 1, 0)^{\mathrm{T}}$，$\boldsymbol{\varepsilon}_3 = (0, 0, 1)^{\mathrm{T}}$ 下的矩阵.

13. 在多项式空间 $F[t]_3$ 中，设 $f(t) = x_1 + x_2 t + x_3 t^2$，定义变换

$$T[f(t)] = (x_2 + x_3) + (x_1 + x_3)t + (x_1 + x_2)t^2$$

(1) 证明 $T$ 是 $F[x]_3$ 的线性变换；

(2) 试求 $T$ 在基 $1, t, t^2$ 下的矩阵；

(3) 试求 $F[x]_3$ 的一组基，使 $T$ 在该基下的矩阵为对角矩阵.

14. 设 $T$ 是 $\mathbf{R}^3$ 中的线性变换 $\boldsymbol{\varepsilon}_1 = (0, 0, 1)^{\mathrm{T}}$，$\boldsymbol{\varepsilon}_2 = (0, 1, 0)^{\mathrm{T}}$，$\boldsymbol{\varepsilon}_3 = (0, 0, T)^{\mathrm{T}}$ 和 $\boldsymbol{\varepsilon}_1' = (1, 1, 1)^{\mathrm{T}}$，$\boldsymbol{\varepsilon}_2' = (1, 1, 0)^{\mathrm{T}}$，$\boldsymbol{\varepsilon}_3' = (1, 0, 0)^{\mathrm{T}}$ 是 $\mathbf{R}^3$ 的两组基，且

$$T(\boldsymbol{\varepsilon}_1') = (1, 0, 0)^{\mathrm{T}}, \quad T(\boldsymbol{\varepsilon}_2') = (2, 1, 0)^{\mathrm{T}}, \quad T(\boldsymbol{\varepsilon}_3') = (3, 2, 1)^{\mathrm{T}}$$

求 $T$ 在基 $\boldsymbol{\varepsilon}_1$，$\boldsymbol{\varepsilon}_2$，$\boldsymbol{\varepsilon}_3$ 下的矩阵.

# 第2章 内积空间

第 1 章介绍的线性空间中向量的基本运算仅是线性运算,并未考虑向量的长度、向量之间的夹角等几何性质. 本章的任务就是在抽象的线性空间中引入这些概念,并得出相应的结果.

## 2.1 内积与欧氏空间

我们先来回顾在几何空间 $\mathbf{R}^3$ 中,向量的长度与向量间的夹角是如何定义的.

设 $\boldsymbol{\alpha}=(a_1, a_2, a_3)^{\mathrm{T}}$, $\boldsymbol{\beta}=(b_1, b_2, b_3)^{\mathrm{T}} \in \mathbf{R}^3$, 称实数

$$\boldsymbol{\alpha}\cdot\boldsymbol{\beta}=a_1b_1+a_2b_2+a_3b_3$$

为向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的数量积.

由数量积,我们将非负实数 $|\boldsymbol{\alpha}|=\sqrt{\boldsymbol{\alpha}\cdot\boldsymbol{\alpha}}$ 称为向量 $\boldsymbol{\alpha}$ 的长度(或模). 当 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 都不是零向量时,$\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 夹角 $\theta$ 的余弦由

$$\cos\theta=\frac{\boldsymbol{\alpha}\cdot\boldsymbol{\beta}}{|\boldsymbol{\alpha}|\cdot|\boldsymbol{\beta}|}$$

给定.

因而可以看到,在几何空间 $\mathbf{R}^3$ 中,向量的长度与向量间的夹角的概念实际上是由所谓的数量积所确定的. 所以,要在一般线性空间中赋予度量,必须先将数量积的概念推广.

**定义 2.1.1** 设 $V$ 是实数域 $\mathbf{R}$ 上的线性空间. 如果对 $V$ 中任意两个向量 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 都有一个实数(记为$(\boldsymbol{\alpha}, \boldsymbol{\beta})$)与它们相对应,并且满足下列条件,则实数$(\boldsymbol{\alpha}, \boldsymbol{\beta})$称为向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的内积:

(i) $(\boldsymbol{\alpha}, \boldsymbol{\beta})=(\boldsymbol{\beta}, \boldsymbol{\alpha})$;

(ii) $(\lambda\boldsymbol{\alpha}, \boldsymbol{\beta})=\lambda(\boldsymbol{\alpha}, \boldsymbol{\beta})$, $\lambda\in\mathbf{R}$;

(iii) $(\boldsymbol{\alpha}+\boldsymbol{\beta}, \boldsymbol{\gamma})=(\boldsymbol{\alpha}, \boldsymbol{\gamma})+(\boldsymbol{\beta}, \boldsymbol{\gamma})$, $\boldsymbol{\gamma}\in V$;

(iv) $(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \geqslant 0$，当且仅当 $\boldsymbol{\alpha} = \mathbf{0}$ 时，等号成立. 而线性空间 $V$ 则称为实内积空间，或欧几里得(Euclid)空间，简称为欧氏空间.

由于向量的内积与向量的线性运算是彼此无关的运算，所以不论内积如何规定，都不会改变线性空间的维数. 故内积空间中的基、维数、坐标等一些相关概念与线性空间类似. 显然，内积空间的子空间仍是欧氏空间.

**例 2.1.1** 在 $\mathbf{R}^n$ 中，对任意的

$$\boldsymbol{\alpha} = (a_1, a_2, \cdots, a_n)^{\mathrm{T}}, \qquad \boldsymbol{\beta} = (b_1, b_2, \cdots, b_n)^{\mathrm{T}}$$

定义 $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n$. 容易验证$(\boldsymbol{\alpha}, \boldsymbol{\beta})$满足内积定义中的 4 条性质，所以是内积，称为 $\mathbf{R}^n$ 的标准内积. 这样 $\mathbf{R}^n$ 就成为欧氏空间，仍记为 $\mathbf{R}^n$. 当 $n = 3$ 时，$(\boldsymbol{\alpha}, \boldsymbol{\beta})$就是通常的数量积.

**例 2.1.2** 考虑 $n^2$ 维线性空间 $\mathbf{R}^{n \times n}$. 如果对任何 $\boldsymbol{A}, \boldsymbol{B} \in \mathbf{R}^{n \times n}$，定义

$$(\boldsymbol{A}, \boldsymbol{B}) = \sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij} b_{ij}$$

则容易验证$(\boldsymbol{A}, \boldsymbol{B})$满足内积定义中的 4 条性质，故是 $\mathbf{R}^{n \times n}$ 的一个内积.

**例 2.1.3** 设 $\mathbf{R}[a, b]$表示定义在闭区间$[a, b]$上的全体实值连续函数的集合，则 $\mathbf{R}[a, b]$是 $\mathbf{R}$ 上的一个线性空间. 对任意 $f(x), g(x) \in \mathbf{R}[a, b]$ 时，令

$$(f(x), g(x)) = \int_a^b f(x) g(x) \mathrm{d}x$$

容易验证 $\mathbf{R}[a, b]$关于以上内积也成一欧氏空间.

同样地，线性空间 $\mathbf{R}[x]$, $\mathbf{R}[x]_n$ 对于例 2.1.3 中定义的内积也构成欧氏空间.

由定义可以推得内积$(\boldsymbol{\alpha}, \boldsymbol{\beta})$具有下列性质：

(i) $(\boldsymbol{\alpha}, \lambda\boldsymbol{\beta}) = \lambda(\boldsymbol{\alpha}, \boldsymbol{\beta})$；

(ii) $(\boldsymbol{\alpha}, \boldsymbol{\beta} + \boldsymbol{\gamma}) = (\boldsymbol{\alpha}, \boldsymbol{\beta}) + (\boldsymbol{\alpha}, \boldsymbol{\gamma})$；

(iii) $(\boldsymbol{\alpha}, \mathbf{0}) = (\mathbf{0}, \boldsymbol{\beta}) = 0$.

其中(iii)的证明如下：

$$(\boldsymbol{\alpha}, \mathbf{0}) = (\boldsymbol{\alpha}, 0\boldsymbol{\beta}) = 0(\boldsymbol{\alpha}, \boldsymbol{\beta}) = 0$$

由定义 2.1.1 中条件(iv)，有 $(\boldsymbol{\alpha}, \boldsymbol{\alpha}) \geqslant 0$，所以对任意向量 $\boldsymbol{\alpha}$，$\sqrt{(\boldsymbol{\alpha}, \boldsymbol{\alpha})}$是有意义的. 因而可以定义 $\boldsymbol{\alpha}$ 的长度.

**定义 2.1.2** 设 $V$ 是内积空间，对任意 $\boldsymbol{\alpha} \in V$，称

$$\| \boldsymbol{\alpha} \| = \sqrt{(\boldsymbol{\alpha}, \boldsymbol{\alpha})}$$

为 $\boldsymbol{\alpha}$ 的长度.

显然，向量的长度一般是正数，只有零向量的长度才是零. 这样定义的长度符合熟知的性质

$$\|k\boldsymbol{\alpha}\| = |k|\ \|\boldsymbol{\alpha}\|$$

这里 $k \in \mathbf{R}$, $\boldsymbol{\alpha} \in V$. 事实上

$$\|k\boldsymbol{\alpha}\| = \sqrt{(k\boldsymbol{\alpha}, k\boldsymbol{\alpha})} = \sqrt{k^2(\boldsymbol{\alpha}, \boldsymbol{\alpha})} = |k|\ \|\boldsymbol{\alpha}\|$$

长度为 1 的向量称为单位向量. 如果 $\boldsymbol{\alpha} \neq \mathbf{0}$, 有

$$\left\| \frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|} \right\| = \frac{1}{\|\boldsymbol{\alpha}\|} \|\boldsymbol{\alpha}\| = 1$$

则 $\frac{\boldsymbol{\alpha}}{\|\boldsymbol{\alpha}\|}$ 是单位向量. 这种对一非零向量除以这个向量的长度成为单位向量的方法，称为向量的单位化.

关于长度、内积，我们有下列基本定理：

**定理 2.1.1** 设 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 是欧氏空间 $V$ 中的任意两个向量，则必有

$$|(\boldsymbol{\alpha}, \boldsymbol{\beta})| \leqslant \|\boldsymbol{\alpha}\|\ \|\boldsymbol{\beta}\|$$

并且当且仅当 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 线性相关时，等号才成立.

**证** 当 $\boldsymbol{\beta} = \mathbf{0}$ 时，则等式自然成立. 下设 $\boldsymbol{\beta} \neq \mathbf{0}$, 令

$$\boldsymbol{\xi} = \boldsymbol{\alpha} - \frac{(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\|\boldsymbol{\beta}\|^2}\boldsymbol{\beta}$$

则有

$$(\boldsymbol{\xi}, \boldsymbol{\beta}) = \left(\boldsymbol{\alpha} - \frac{(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\|\boldsymbol{\beta}\|^2}\boldsymbol{\beta}, \boldsymbol{\beta}\right) = (\boldsymbol{\alpha}, \boldsymbol{\beta}) - (\boldsymbol{\alpha}, \boldsymbol{\beta}) = 0$$

$$0 \leqslant (\boldsymbol{\xi}, \boldsymbol{\xi}) = \left(\boldsymbol{\xi}, \boldsymbol{\alpha} - \frac{(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\|\boldsymbol{\beta}\|^2}\boldsymbol{\beta}\right) = (\boldsymbol{\xi}, \boldsymbol{\alpha}) = \left(\boldsymbol{\alpha} - \frac{(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\|\boldsymbol{\beta}\|^2}\boldsymbol{\beta}, \boldsymbol{\alpha}\right)$$
$$= \|\boldsymbol{\alpha}\|^2 - \frac{|(\boldsymbol{\alpha}, \boldsymbol{\beta})|^2}{\|\boldsymbol{\beta}\|^2}$$

即
$$|(\boldsymbol{\alpha}, \boldsymbol{\beta})|^2 \leqslant \|\boldsymbol{\alpha}\|^2\ \|\boldsymbol{\beta}\|^2$$

两边开平方便得所要证明的不等式.

当 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 线性相关时，等号显然成立；反之，如果等号成立，则 $\boldsymbol{\xi} = \mathbf{0}$, 即

$$\boldsymbol{\alpha} - \frac{(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\|\boldsymbol{\beta}\|^2}\boldsymbol{\beta} = \mathbf{0}$$

因而 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 线性相关. □

上述定理 2.1.1 中的不等式称为柯西-施瓦茨(Cauchy-Schwarz)不等式,这个不等式有着非常广泛的应用. 对于例 2.1.1 的欧氏空间 $\mathbf{R}^n$,有

$$|a_1b_1+a_2b_2+\cdots+a_nb_n|\leqslant\sqrt{a_1^2+a_2^2+\cdots+a_n^2}\sqrt{b_1^2+b_2^2+\cdots+b_n^2}$$

对于例 2.1.3 的欧氏空间 $\mathbf{R}[a, b]$,有

$$\left|\int_a^b f(x)g(x)\mathrm{d}x\right|\leqslant\left(\int_a^b f^2(x)\mathrm{d}x\right)^{1/2}\left(\int_a^b g^2(x)\mathrm{d}x\right)^{1/2}$$

定理 2.1.1 的一个推论是

$$\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|\leqslant\|\boldsymbol{\alpha}\|+\|\boldsymbol{\beta}\|$$

称为三角不等式. 事实上,由于

$$\begin{aligned}\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|^2&=(\boldsymbol{\alpha}+\boldsymbol{\beta}, \boldsymbol{\alpha}+\boldsymbol{\beta})=(\boldsymbol{\alpha}, \boldsymbol{\alpha})+2(\boldsymbol{\alpha}, \boldsymbol{\beta})+(\boldsymbol{\beta}, \boldsymbol{\beta})\\&\leqslant\|\boldsymbol{\alpha}\|^2+2\|\boldsymbol{\alpha}\|\|\boldsymbol{\beta}\|+\|\boldsymbol{\beta}\|^2\\&=(\|\boldsymbol{\alpha}\|+\|\boldsymbol{\beta}\|)^2\end{aligned}$$

两端开平方即得.

当 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 都不是零向量时,由柯西-施瓦茨不等式可得

$$\frac{|(\boldsymbol{\alpha}, \boldsymbol{\beta})|}{\|\boldsymbol{\alpha}\|\|\boldsymbol{\beta}\|}\leqslant 1$$

亦即

$$-1\leqslant\frac{(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\|\boldsymbol{\alpha}\|\|\boldsymbol{\beta}\|}\leqslant 1$$

因此,我们可以用等式

$$\cos\theta=\frac{(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\|\boldsymbol{\alpha}\|\|\boldsymbol{\beta}\|}$$

来定义两个非零向量的夹角 $\theta$,且限制 $\theta$ 的取值范围为 $0\leqslant\theta\leqslant\pi$. 当 $(\boldsymbol{\alpha}, \boldsymbol{\beta})=0$ 时,称 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 是正交的,且记为 $\boldsymbol{\alpha}\perp\boldsymbol{\beta}$.

**例 2.1.4** 若 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 是两个正交向量,则

$$\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|^2=\|\boldsymbol{\alpha}\|^2+\|\boldsymbol{\beta}\|^2$$

这利用内积性质及正交条件是不难证明的. 这里留给读者作为练习.

## 2.2 欧氏空间的正交基

**定义 2.2.1** 欧氏空间 $V$ 中一组非零的向量,如果它们是两两正交的,就称为一正交向量组.

不难证明,正交向量组是线性无关的. 事实上,设正交向量组 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s$ 有一线性关系

$$k_1\boldsymbol{\alpha}_1 + k_2\boldsymbol{\alpha}_2 + \cdots + k_s\boldsymbol{\alpha}_s = \mathbf{0}$$

用 $\boldsymbol{\alpha}_i$ 与等式两边作内积,即得

$$k_i(\boldsymbol{\alpha}_i, \boldsymbol{\alpha}_i) = 0$$

由 $\boldsymbol{\alpha}_i \neq \mathbf{0}$,有$(\boldsymbol{\alpha}_i, \boldsymbol{\alpha})_i > 0$,从而 $k_i = 0$ $(i = 1, 2, \cdots, s)$. 这就证明了 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s$ 是线性无关的.

**定义 2.2.2** 在 $n$ 维欧氏空间中,由 $n$ 个向量组成的正交向量组称为正交基. 如果正交基中每个基向量的长度都等于 1,则这组正交基称为标准正交基.

简单地,若 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$ 是 $n$ 维欧氏空间 $V$ 的一组非零向量,且满足条件

$$(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_j) = \begin{cases} 1, & \text{当 } i = j \\ 0, & \text{当 } i \neq j \end{cases}$$

$(i, j = 1, 2, \cdots, n)$,则 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$ 即为一标准正交基.

设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 是 $n$ 维欧氏空间 $V$ 的一组基,要求 $V$ 的一组标准正交基. 这也就是要找一组两两正交的单位向量 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$,使 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$ 与 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 等价. 这样一个问题,称为把 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 这组基标准正交化.

我们可以用下列办法把 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 标准正交化: 取

$$\boldsymbol{\beta}_1 = \boldsymbol{\alpha}_1$$

$$\boldsymbol{\beta}_2 = \boldsymbol{\alpha}_2 - \frac{(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_1)}{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_1)}\boldsymbol{\beta}_1$$

$$\cdots\cdots$$

$$\boldsymbol{\beta}_n = \boldsymbol{\alpha}_n - \frac{(\boldsymbol{\alpha}_n, \boldsymbol{\beta}_1)}{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_1)}\boldsymbol{\beta}_1 - \cdots - \frac{(\boldsymbol{\alpha}_n, \boldsymbol{\beta}_{n-1})}{(\boldsymbol{\beta}_{n-1}, \boldsymbol{\beta}_{n-1})}\boldsymbol{\beta}_{n-1}$$

容易验证 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n$ 两两正交,且 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n$ 与 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 等价.

然后只要把它们单位化,即取

$$\boldsymbol{\gamma}_1 = \frac{1}{\|\boldsymbol{\beta}_1\|}\boldsymbol{\beta}_1, \boldsymbol{\gamma}_2 = \frac{1}{\|\boldsymbol{\beta}_2\|}\boldsymbol{\beta}_2, \cdots, \boldsymbol{\gamma}_n = \frac{1}{\|\boldsymbol{\beta}_n\|}\boldsymbol{\beta}_n$$

就是 $V$ 的一个标准正交基.

上述从线性无关向量组 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 导出正交向量组 $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n$,进而得到标准正交向量组 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$ 的过程称为施密特(Schimidt)正交化方法. 由此可得以下定理.

**定理 2.2.1** 任一 $n$ 维欧氏空间 $V$ 都存在标准正交基.

**推论 2.2.1** 设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 是 $n$ 维欧氏空间 $V$ 的一组基,则在 $V$ 中存在标准正交基 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$,使得

$$(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n) = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n)\boldsymbol{R}$$

其中 $\boldsymbol{R}$ 是主对角元为正数的上三角矩阵.

**证** 由上面的施密特正交化过程,有等式

$$\begin{cases} \boldsymbol{\alpha}_1 = \boldsymbol{\beta}_1 \\ \boldsymbol{\alpha}_2 = \dfrac{(\boldsymbol{\alpha}_2, \boldsymbol{\beta}_1)}{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_1)}\boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 \\ \cdots\cdots \\ \boldsymbol{\alpha}_n = \dfrac{(\boldsymbol{\alpha}_n, \boldsymbol{\beta}_1)}{(\boldsymbol{\beta}_1, \boldsymbol{\beta}_1)}\boldsymbol{\beta}_1 + \dfrac{(\boldsymbol{\alpha}_n, \boldsymbol{\beta}_2)}{(\boldsymbol{\beta}_2, \boldsymbol{\beta}_2)}\boldsymbol{\beta}_2 + \cdots + \dfrac{(\boldsymbol{\alpha}_n, \boldsymbol{\beta}_{n-1})}{(\boldsymbol{\beta}_{n-1}, \boldsymbol{\beta}_{n-1})}\boldsymbol{\beta}_{n-1} + \boldsymbol{\beta}_n \end{cases}$$

再由 $\boldsymbol{\gamma}_i = \dfrac{1}{\|\boldsymbol{\beta}_i\|}\boldsymbol{\beta}_i (i = 1, 2, \cdots, n)$ 有

$$\boldsymbol{\beta}_i = \|\boldsymbol{\beta}_i\| \boldsymbol{\gamma}_i, \quad i = 1, 2, \cdots, n$$

代入得

$$\begin{cases} \boldsymbol{\alpha}_1 = \|\boldsymbol{\beta}_1\| \boldsymbol{\gamma}_1 \\ \boldsymbol{\alpha}_2 = (\boldsymbol{\alpha}_2, \boldsymbol{\gamma}_1)\boldsymbol{\gamma}_1 + \|\boldsymbol{\beta}_2\| \boldsymbol{\gamma}_2 \\ \cdots\cdots \\ \boldsymbol{\alpha}_n = (\boldsymbol{\alpha}_n, \boldsymbol{\gamma}_1)\boldsymbol{\gamma}_1 + (\boldsymbol{\alpha}_n, \boldsymbol{\gamma}_2)\boldsymbol{\gamma}_2 + \cdots + (\boldsymbol{\alpha}_n, \boldsymbol{\gamma}_{n-1})\boldsymbol{\gamma}_{n-1} + \|\boldsymbol{\beta}_n\| \boldsymbol{\gamma}_n \end{cases}$$

所以有

$$(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n) = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n)\begin{pmatrix} \|\boldsymbol{\beta}_1\| & (\boldsymbol{\alpha}_2, \boldsymbol{\gamma}_1) & \cdots & (\boldsymbol{\alpha}_n, \boldsymbol{\gamma}_1) \\ & \|\boldsymbol{\beta}_2\| & \cdots & (\boldsymbol{\alpha}_n, \boldsymbol{\gamma}_2) \\ & & \ddots & \vdots \\ & & & \|\boldsymbol{\beta}_n\| \end{pmatrix}$$

$$= (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n)\boldsymbol{R}$$ □

为什么在欧氏空间中总是取标准正交基，而不用一般的基呢？下面的一些性质可以看出一些理由.

**定理 2.2.2** 设 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$ 是 $n$ 维欧氏空间 $V$ 中的一组标准正交基，对任意 $\boldsymbol{\alpha} \in V$, $\boldsymbol{\alpha} = x_1\boldsymbol{\gamma}_1 + x_2\boldsymbol{\gamma}_2 + \cdots + x_n\boldsymbol{\gamma}_n$，则

$$x_i = (\boldsymbol{\alpha}, \boldsymbol{\gamma}_i), \quad i = 1, 2, \cdots, n$$

**证** 设 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$ 是 $V$ 的一组标准正交基，对任意 $\boldsymbol{\alpha} \in V$, $\boldsymbol{\alpha} = x_1\boldsymbol{\gamma}_1 + x_2\boldsymbol{\gamma}_2 + \cdots + x_n\boldsymbol{\gamma}_n$ 用 $\boldsymbol{\gamma}_i$ 在等式两边分别作内积，有

$$\begin{aligned}(\boldsymbol{\alpha}, \boldsymbol{\gamma}_i) &= (x_1\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_i) + \cdots + (x_i\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_i) + \cdots + (x_n\boldsymbol{\gamma}_n, \boldsymbol{\gamma}_i) \\ &= x_i(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_i), \quad i = 1, 2, \cdots, n\end{aligned}$$

由 $(\boldsymbol{\gamma}_i, \boldsymbol{\gamma}_i) = 1$ 知

$$x_i = (\boldsymbol{\alpha}, \boldsymbol{\gamma}_i), \quad i = 1, 2, \cdots, n$$ □

**定理 2.2.3** 设 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$ 为 $n$ 维欧氏空间 $V$ 的一组标准正交基，对任意 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in V$，若

$$\boldsymbol{\alpha} = x_1\boldsymbol{\gamma}_1 + x_2\boldsymbol{\gamma}_2 + \cdots + x_n\boldsymbol{\gamma}_n$$

$$\boldsymbol{\beta} = y_1\boldsymbol{\gamma}_1 + y_2\boldsymbol{\gamma}_2 + \cdots + y_n\boldsymbol{\gamma}_n$$

则

$$(\boldsymbol{\alpha}, \boldsymbol{\beta}) = x_1y_1 + x_2y_2 + \cdots + x_ny_n$$

**证** 由内积性质及标准正交基的定义可得. □

由此可见，在标准正交基下，有限维欧氏空间向量内积可以由坐标的一个简单表达式来描述.

**定理 2.2.4** 设 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$; $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n$ 都是 $n$ 维欧氏空间 $V$ 的标准正交基，且有

$$(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n) = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n)\boldsymbol{A}$$

则 $\boldsymbol{A}$ 是正交矩阵.

**证** 设 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$ 及 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n$ 是 $n$ 维欧氏空间 $V$ 的两组标准正交基，它们之间的过渡矩阵为 $\boldsymbol{A} = (a_{ij})$，即

$$(\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n) = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n)\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

因为 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n$ 是标准正交基，所以

$$(\boldsymbol{\eta}_i, \boldsymbol{\eta}_j) = \begin{cases} 1, & \text{当 } i = j \\ 0, & \text{当 } i \neq j \end{cases}$$

矩阵 $\boldsymbol{A}$ 的各列就是 $\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \cdots, \boldsymbol{\eta}_n$ 在标准正交基 $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$ 下的坐标，由定理 2.2.3 有

$$a_{1i}a_{1j} + a_{2i}a_{2j} + \cdots + a_{ni}a_{nj} = \begin{cases} 1, & \text{当 } i = j \\ 0, & \text{当 } i \neq j \end{cases}$$

这组关系式表明矩阵 $\boldsymbol{A}$ 满足等式

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{E}$$

其中 $\boldsymbol{A}^{\mathrm{T}}$ 是 $\boldsymbol{A}$ 的转置矩阵. 可知 $\boldsymbol{A}$ 为正交矩阵. □

## 2.3 欧氏空间的同构

所有的有限维线性空间可以按其维数进行分类，对内积空间来说也有相同的分类结果，本节我们研究这一问题.

**定义 2.3.1** 设 $V$ 与 $W$ 是欧氏空间，如果存在一个线性空间的同构 $T: V \to W$ 使得对任意的 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in V$，有

$$(T(\boldsymbol{\alpha}), T(\boldsymbol{\beta})) = (\boldsymbol{\alpha}, \boldsymbol{\beta})$$

则称 $T$ 是内积同构映射，而 $V$ 与 $W$ 称为是同构的.

内积同构映射除了要保持线性空间的线性运算外，还必须是保持内积运算的双射.

内积同构关系显然也满足自反性、对称性和传递性.

与线性空间一样，我们有下述定理.

**定理 2.3.1** 设 $V$ 与 $W$ 是有限维欧氏空间，则 $V$ 与 $W$ 是同构的，当且仅当 $\dim V = \dim W$.

**证** 如果 $V$ 与 $W$ 是同构的，则它们也是同构的线性空间，故由线性空间的同构知，$\dim V = \dim W$. 必要性成立.

反之，设 $\dim V = \dim W$. 在 $V$ 及 $W$ 中分别取标准正交基 $\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n\}$，$\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n\}$. 定义映射 $T: V \to W$ 为

$$T(\boldsymbol{\alpha}) = T(x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n) = x_1\boldsymbol{\beta}_1 + x_2\boldsymbol{\beta}_2 + \cdots + x_n\boldsymbol{\beta}_n$$

其中 $\boldsymbol{\alpha}=\sum_{i=1}^{n}x_i\boldsymbol{\alpha}_i$ 为 $V$ 中任一元. 易证, $T$ 是线性空间 $V$ 与 $W$ 的一个同构映射.

我们证明 $T$ 是保持内积运算的. 任取 $\boldsymbol{\beta}=y_1\boldsymbol{\alpha}_1+y_2\boldsymbol{\alpha}_2+\cdots+y_n\boldsymbol{\alpha}_n\in V$, 则

$$(T(\boldsymbol{\alpha}),\ T(\boldsymbol{\beta}))=\left(\sum_{i=1}^{n}x_i\boldsymbol{\beta}_i,\ \sum_{j=1}^{n}y_j\boldsymbol{\beta}_j\right)=\sum_{i=1}^{n}x_iy_i=(\boldsymbol{\alpha},\ \boldsymbol{\beta})$$

所以, $V$ 与 $W$ 是内积同构的. □

**定理 2.3.2** 设 $V$, $W$ 是有限维欧氏空间,则 $V$ 与 $W$ 是内积同构的充要条件是存在映射 $T: V\to W$ 将 $V$ 的标准正交基映射成 $W$ 的标准正交基.

**证** 设 $V$ 与 $W$ 是内积同构的, $\{\boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2,\ \cdots,\ \boldsymbol{\varepsilon}_n\}$ 是 $V$ 的一个标准正交基,则存在内积同构映射 $T: V\to W$. 由于 $T$ 也是线性空间的同构映射,可知 $\{T(\boldsymbol{\varepsilon}_1),\ T(\boldsymbol{\varepsilon}_2),\ \cdots,\ T(\boldsymbol{\varepsilon}_n)\}$ 必是 $W$ 的基. 我们证明 $\{T(\boldsymbol{\varepsilon}_1),\ T(\boldsymbol{\varepsilon}_2),\ \cdots,\ T(\boldsymbol{\varepsilon}_n)\}$ 是标准正交的. 事实上,有

$$(T(\boldsymbol{\varepsilon}_i),\ T(\boldsymbol{\varepsilon}_j))=(\boldsymbol{\varepsilon}_i,\ \boldsymbol{\varepsilon}_j)=\delta_{ij},\quad i,\ j=1,\ 2,\ \cdots,\ n$$

所以,必要性是成立的.

反之,假设有映射 $T: V\to W$ 将 $V$ 的标准正交基 $\{\boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2,\ \cdots,\ \boldsymbol{\varepsilon}_n\}$ 映射成 $W$ 的标准正交基 $\{T(\boldsymbol{\varepsilon}_1),\ T(\boldsymbol{\varepsilon}_2),\ \cdots,\ T(\boldsymbol{\varepsilon}_n)\}$. 可知 $T$ 是线性空间的同构,因而我们只需证明 $T$ 是保持内积运算的.

任取 $\boldsymbol{\alpha}=x_1\boldsymbol{\varepsilon}_1+x_2\boldsymbol{\varepsilon}_2+\cdots+x_n\boldsymbol{\varepsilon}_n$, $\boldsymbol{\beta}=y_1\boldsymbol{\varepsilon}_1+y_2\boldsymbol{\varepsilon}_2+\cdots+y_n\boldsymbol{\varepsilon}_n\in V$, 且

$$\begin{aligned}(T(\boldsymbol{\alpha}),\ T(\boldsymbol{\beta}))&=\left(\sum_{i=1}^{n}x_iT(\boldsymbol{\varepsilon}_i),\ \sum_{j=1}^{n}y_jT(\boldsymbol{\varepsilon}_j)\right)=\sum_{i,\ j=1}^{n}x_iy_j(T(\boldsymbol{\varepsilon}_i),\ T(\boldsymbol{\varepsilon}_j))\\&=\sum_{i=1}^{n}x_iy_i=(\boldsymbol{\alpha},\ \boldsymbol{\beta})\end{aligned}$$

这就证明了充分性. □

由定理 2.3.2 知,所有 $n$ 维欧氏空间 $V$ 均与 $\mathbf{R}^n$ 同构.

## 2.4 正交补

**定义 2.4.1** 设 $W_1$, $W_2$ 是欧氏空间的两个子集,如果对任意 $\boldsymbol{\alpha}\in W_1$, $\boldsymbol{\beta}\in W_2$, 有 $(\boldsymbol{\alpha},\ \boldsymbol{\beta})=0$, 则称集合 $W_1$ 与 $W_2$ 是相互正交的,记为 $W_1\perp W_2$.

如果 $W_1$, $W_2$ 是有限维的非平凡子空间,则容易证明 $W_1\perp W_2$, 当且仅当它们的基是相互正交的.

一个向量与一个子空间正交的充要条件是该向量与子空间的基正交.

**例 2.4.1** 设 $W$ 是欧氏空间 $V$ 的一个子集. 证明 $V$ 中所有与 $W$ 正交的向量所成之集是 $V$ 的子空间.

**证** 记 $W^{\perp}=\{\boldsymbol{\beta}\mid\{\boldsymbol{\beta}\}\perp W\}$. 现任取 $\boldsymbol{\beta}_1,\ \boldsymbol{\beta}_2\in W^{\perp}$, $\lambda_1,\ \lambda_2$ 为数,则对任意 $\boldsymbol{\alpha}\in W$, 有

$$(\lambda_1\boldsymbol{\beta}_1+\lambda_2\boldsymbol{\beta}_2,\ \boldsymbol{\alpha})=\lambda_1(\boldsymbol{\beta}_1,\ \boldsymbol{\alpha})+\lambda_2(\boldsymbol{\beta}_2,\ \boldsymbol{\alpha})=0$$

故 $\lambda_1\boldsymbol{\beta}_1+\lambda_2\boldsymbol{\beta}_2\in W^{\perp}$. 这就证明了 $W^{\perp}$是子空间.

**定义 2.4.2** $W^{\perp}$称为 $W$ 的正交补.

**定理 2.4.1** 设 $W$ 是欧氏空间 $V$ 的有限维子空间,则

$$V=W\oplus W^{\perp}$$

**证** 取 $W$ 的标准正交基$\{\boldsymbol{\varepsilon}_1,\ \boldsymbol{\varepsilon}_2,\ \cdots,\ \boldsymbol{\varepsilon}_m\}$,对任意的 $\boldsymbol{\alpha}\in V$, 令

$$\boldsymbol{\alpha}'=(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_1)\boldsymbol{\varepsilon}_1+(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_2)\boldsymbol{\varepsilon}_2+\cdots+(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_m)\boldsymbol{\varepsilon}_m$$

$$\boldsymbol{\alpha}''=\boldsymbol{\alpha}-\boldsymbol{\alpha}'$$

则由定义可知, $\boldsymbol{\alpha}'\in W$. 下证 $\boldsymbol{\alpha}''\in W^{\perp}$. 事实上,对 $j=1,\ 2,\ \cdots,\ m$,有

$$\begin{aligned}(\boldsymbol{\alpha}'',\ \boldsymbol{\varepsilon}_j)&=(\boldsymbol{\alpha}-\boldsymbol{\alpha}',\ \boldsymbol{\varepsilon}_j)=(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_j)-(\boldsymbol{\alpha}',\ \boldsymbol{\varepsilon}_j)\\&=(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_j)-\left(\sum_{i=1}^{m}(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_i)\boldsymbol{\varepsilon}_i,\ \boldsymbol{\varepsilon}_j\right)\\&=(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_j)-\sum_{i=1}^{m}(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_i)(\boldsymbol{\varepsilon}_i,\ \boldsymbol{\varepsilon}_j)\\&=(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_j)-(\boldsymbol{\alpha},\ \boldsymbol{\varepsilon}_j)=0\end{aligned}$$

所以

$$\boldsymbol{\alpha}=\boldsymbol{\alpha}'+\boldsymbol{\alpha}''\in W+W^{\perp}$$

此时

$$V=W+W^{\perp}$$

若存在 $\boldsymbol{\beta}\in W\cap W^{\perp}$, 则必有

$$(\boldsymbol{\beta},\ \boldsymbol{\beta})=0$$

从而

$$\boldsymbol{\beta}=\mathbf{0}$$

故

$$V=W\oplus W^{\perp}$$

□

**推论 2.4.1** 若 $V$ 是有限维欧氏空间, $W$ 是 $V$ 的子空间,则 $W=W^{\perp\perp}$.

由定理 2.4.1 知道,对每个 $V$ 中的向量 $\boldsymbol{\alpha}$,必可以找到 $W$ 中的唯一向量 $\boldsymbol{\alpha}'$及 $W^{\perp}$ 中的唯一向量 $\boldsymbol{\alpha}''$,使 $\boldsymbol{\alpha}=\boldsymbol{\alpha}'+\boldsymbol{\alpha}''$, 我们称 $\boldsymbol{\alpha}'$为 $\boldsymbol{\alpha}$ 在子空间 $W$ 上的正交投影, $V$

中的向量在其子空间 $W$ 上的正交投影具有重要的几何意义.

**定理 2.4.2** 设 $W$ 是欧氏空间 $V$ 的子空间, $\boldsymbol{\alpha}\in V$. 若 $\boldsymbol{\alpha}_0$ 是 $\boldsymbol{\alpha}$ 在 $W$ 上的正交投影,则

$$\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| = \min_{\boldsymbol{\beta}\in W}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|$$

且使上式成立的 $\boldsymbol{\alpha}_0$ 是唯一的. $\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|$ 称为向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\alpha}_0$ 的距离.

**证** 我们要证明的是定义在 $W$ 上的非负实值函数 $\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|$. $\boldsymbol{\beta}\in W$ 在 $W$ 上可以取得唯一的最小值点 $\boldsymbol{\alpha}_0$.

由正交投影的定义可知, $\boldsymbol{\alpha}_0\in W$ 且 $\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\in W^{\perp}$. 对任意的 $\boldsymbol{\beta}\in W$, 因

$$\boldsymbol{\alpha}-\boldsymbol{\beta}=(\boldsymbol{\alpha}-\boldsymbol{\alpha}_0)+(\boldsymbol{\alpha}_0-\boldsymbol{\beta}),\quad \boldsymbol{\alpha}_0-\boldsymbol{\beta}\in W$$

故 $\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\perp\boldsymbol{\alpha}_0-\boldsymbol{\beta}$. 从而

$$\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|^2 = \|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|^2+\|\boldsymbol{\alpha}_0-\boldsymbol{\beta}\|^2\geqslant\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|^2 \tag{2.4.1}$$

这说明函数 $\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|$ 有下界 $\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|$. 显然取 $\boldsymbol{\beta}=\boldsymbol{\alpha}_0$ 时,该函数可以达到这个下界,所以

$$\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| = \min_{\boldsymbol{\beta}\in W}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|$$

如果还有 $\boldsymbol{\alpha}_0'\in W$, 使

$$\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0'\| = \|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\| = \min_{\boldsymbol{\beta}\in W}\|\boldsymbol{\alpha}-\boldsymbol{\beta}\|$$

则由式(2.4.1)知

$$\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0'\|^2 = \|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|^2+\|\boldsymbol{\alpha}_0-\boldsymbol{\alpha}_0'\|^2 = \|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|^2$$

从而有

$$\|\boldsymbol{\alpha}_0-\boldsymbol{\alpha}_0'\| = 0$$

于是

$$\boldsymbol{\alpha}_0=\boldsymbol{\alpha}_0'$$

□

定理 2.4.2 说明,向量 $\boldsymbol{\alpha}$ 到 $W$ 的各向量间的距离,以垂线 $\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_0\|$ 最短.

**例 2.4.2** 证明 $\mathbf{R}^3$ 中定点 $(x_0, y_0, z_0)$ 到已知平面 $Ax+By+Cz=0$ 的距离等于

$$\frac{|Ax_0+By_0+Cz_0|}{\sqrt{A^2+B^2+C^2}}$$

**证** 平面上的所有向量构成之集是 $\mathbf{R}^3$ 的子空间,记为 $W$. 根据定理 2.4.2,设向量 $\boldsymbol{\alpha}=(x_0, y_0, z_0)$ 到 $W$ 的正交投影为 $\boldsymbol{\alpha}'$ 时, $\boldsymbol{\alpha}$ 到 $W$ 的距离应为 $\|\boldsymbol{\alpha}-\boldsymbol{\alpha}'\|$. 但

$\mathbf{R}^3 = W \oplus W^{\perp}$,因此有 $\boldsymbol{\alpha}'' \in W^{\perp}$,使 $\boldsymbol{\alpha} = \boldsymbol{\alpha}' + \boldsymbol{\alpha}''$, 从而

$$\| \boldsymbol{\alpha}'' \| = \| \boldsymbol{\alpha} - \boldsymbol{\alpha}' \|$$

所以只要求出 $\boldsymbol{\alpha}$ 在 $W^{\perp}$ 上的正交投影 $\boldsymbol{\alpha}''$,则由上式即可以得出 $\boldsymbol{\alpha}$ 到 $W$ 的距离.

取 $\boldsymbol{e} = \frac{1}{\sqrt{A^2+B^2+C^2}}(A, B, C)$, 则$\{\boldsymbol{e}\}$是 $W^{\perp}$ 的标准正交基,所以, $\boldsymbol{\alpha}'' = (\boldsymbol{\alpha}, \boldsymbol{e})\boldsymbol{e}$, 从而

$$\| \boldsymbol{\alpha}'' \| = | (\boldsymbol{\alpha}, \boldsymbol{e}) | = \frac{| Ax_0 + By_0 + Cz_0 |}{\sqrt{A^2+B^2+C^2}}$$

定理 2.4.2 所阐述的简单的几何事实可以用来解决一些实际问题. 其中的一个应用是解决最小二乘法问题. 下面我们用欧氏空间的概念来表达最小二乘法,并给出最小二乘解所满足的代数条件.

设有方程组

$$\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$$

这里 $\boldsymbol{A} = (a_{ij})_{n\times s}$, $\boldsymbol{b} = (b_1, b_2, \cdots, b_n)^{\mathrm{T}}$, $\boldsymbol{x} = (x_1, x_2, \cdots, x_s)^{\mathrm{T}}$ ($x_i$ 为实变数). 如果方程组无解,则称之为不相容. 此时,只好设法找出一组数 $x_1^0, x_2^0, \cdots, x_s^0$,使平方偏差

$$\delta = \sum_{i=1}^{n} (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{is}x_s - b_i)^2$$

最小. 这组数称为该方程组的最小二乘解,这种方法称为最小二乘法.

令 $\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x}$, 则 $\boldsymbol{y}$ 当然是一个 $n$ 维列向量. 上述平方偏差 $\delta$ 也就是 $\| \boldsymbol{y} - \boldsymbol{b} \|^2$, 而最小二乘法就是要找出一组数 $x_1^0, x_2^0, \cdots, x_s^0$,使 $\boldsymbol{y}$ 与 $\boldsymbol{b}$ 的距离最小.

设 $\boldsymbol{A} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s)$, $\boldsymbol{\alpha}_i$ 表示 $\boldsymbol{A}$ 的第 $i$ 列,则有

$$\boldsymbol{y} = x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_s\boldsymbol{\alpha}_s \tag{2.4.2}$$

显然 $\boldsymbol{y} \in \mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$, 故最小二乘法可以叙述为: 求 $\boldsymbol{x}$ 使 $\| \boldsymbol{y} - \boldsymbol{b} \|^2$ 最小,即在 $\mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$中找出一向量 $\boldsymbol{y}$,使得向量 $\boldsymbol{b}$ 到 $\boldsymbol{y}$ 的距离比到子空间 $\mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$中其他向量的距离都短.

由定理 2.4.2 知,若式 2.4.2 中 $\boldsymbol{y}$ 为所求的向量,则

$$\boldsymbol{c} = \boldsymbol{b} - \boldsymbol{y} = \boldsymbol{b} - \boldsymbol{A}\boldsymbol{x}$$

必须垂直于子空间 $\mathrm{Span}\{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_s\}$,为此有

$$(\boldsymbol{c}, \boldsymbol{\alpha}_1) = (\boldsymbol{c}, \boldsymbol{\alpha}_2) = \cdots = (\boldsymbol{c}, \boldsymbol{\alpha}_s) = 0$$

这条件相当于

$$\boldsymbol{\alpha}_1^{\mathrm{T}}\boldsymbol{c}=0,\ \boldsymbol{\alpha}_2^{\mathrm{T}}\boldsymbol{c}=0,\ \cdots,\ \boldsymbol{\alpha}_s^{\mathrm{T}}\boldsymbol{c}=0$$

这组等式相当于

$$\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{b}-\boldsymbol{A}\boldsymbol{x})=\boldsymbol{0}$$

即

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$$

这就是最小二乘解所满足的代数方程.

**例 2.4.3** 设

$$\boldsymbol{A}=\begin{pmatrix}1&1&1\\2&4&1\\1&3&0\end{pmatrix},\qquad \boldsymbol{b}=\begin{pmatrix}1\\1\\1\end{pmatrix}$$

求方程组 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 的全部最小二乘解.

**解** 易知 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 是不相容方程组,且

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\begin{pmatrix}6&12&3\\12&26&5\\3&5&2\end{pmatrix},\qquad \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}=\begin{pmatrix}4\\8\\2\end{pmatrix}$$

求解 $\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}$ 是不相容方程组 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}$:

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A},\ \boldsymbol{A}^{\mathrm{T}}\boldsymbol{b})\rightarrow\cdots\rightarrow\begin{pmatrix}1&0&\dfrac{3}{2}&\dfrac{2}{3}\\0&1&-\dfrac{1}{2}&0\\0&0&0&0\end{pmatrix}$$

得全部解

$$\boldsymbol{x}=k\begin{pmatrix}-3\\1\\0\end{pmatrix}+\begin{pmatrix}\dfrac{2}{3}\\0\\0\end{pmatrix},\quad k\in\mathbf{R}$$

## 2.5 正交变换

本节在欧氏空间上讨论一种特别的线性变换——正交变换.

**定义 2.5.1** 设 $T$ 是欧氏空间 $V$ 的线性变换，若 $T$ 保持 $V$ 中的向量内积不变，即对任何的 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in V$，都有

$$(T\boldsymbol{\alpha}, T\boldsymbol{\beta}) = (\boldsymbol{\alpha}, \boldsymbol{\beta})$$

则称 $T$ 为 $V$ 上的一个正交变换.

正交变换可以从几个不同的方面来加以描述.

**定理 2.5.1** 设 $T$ 是 $n$ 维欧氏空间 $V$ 的一个线性变换，则下列各个命题是等价的：

(i) $T$ 是正交变换；

(ii) $T$ 保持向量的长度不变，即对任一 $\boldsymbol{\alpha} \in V$，都有 $\| T\boldsymbol{\alpha} \| = \| \boldsymbol{\alpha} \|$；

(iii) 若 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 是 $V$ 的一组标准正交基，则 $T\boldsymbol{e}_1, T\boldsymbol{e}_2, \cdots, T\boldsymbol{e}_n$ 也是 $V$ 的一组标准正交基；

(iv) $T$ 在 $V$ 的任一标准正交基下的矩阵是正交矩阵.

**证** (i)⇔(ii)

若 $T$ 是正交变换，则由定义知

$$(T\boldsymbol{\alpha}, T\boldsymbol{\alpha}) = (\boldsymbol{\alpha}, \boldsymbol{\alpha})$$

两边开平方得

$$\| T\boldsymbol{\alpha} \| = \| \boldsymbol{\alpha} \|$$

反之，若 $T$ 保持向量的长度不变，即

$$\| T\boldsymbol{\alpha} \| = \sqrt{(T\boldsymbol{\alpha}, T\boldsymbol{\alpha})} = \| \boldsymbol{\alpha} \| = \sqrt{(\boldsymbol{\alpha}, \boldsymbol{\alpha})}$$

$$\| T\boldsymbol{\beta} \| = \sqrt{(T\boldsymbol{\beta}, T\boldsymbol{\beta})} = \| \boldsymbol{\beta} \| = \sqrt{(\boldsymbol{\beta}, \boldsymbol{\beta})}$$

则

$$(T\boldsymbol{\alpha}, T\boldsymbol{\alpha}) = (\boldsymbol{\alpha}, \boldsymbol{\alpha})$$

$$(T\boldsymbol{\beta}, T\boldsymbol{\beta}) = (\boldsymbol{\beta}, \boldsymbol{\beta})$$

同样有

$$(T(\boldsymbol{\alpha}+\boldsymbol{\beta}), T(\boldsymbol{\alpha}+\boldsymbol{\beta})) = (\boldsymbol{\alpha}+\boldsymbol{\beta}, \boldsymbol{\alpha}+\boldsymbol{\beta})$$

将上式两边展开即得

$$(T\boldsymbol{\alpha}, T\boldsymbol{\alpha}) + 2(T\boldsymbol{\alpha}, T\boldsymbol{\beta}) + (T\boldsymbol{\beta}, T\boldsymbol{\beta}) = (\boldsymbol{\alpha}, \boldsymbol{\alpha}) + 2(\boldsymbol{\alpha}, \boldsymbol{\beta}) + (\boldsymbol{\beta}, \boldsymbol{\beta})$$

从而有

$$(T\boldsymbol{\alpha}, T\boldsymbol{\beta}) = (\boldsymbol{\alpha}, \boldsymbol{\beta})$$

即 $T$ 为正交变换.

(i)⇔(iii)

设 $T$ 是正交变换,则对 $V$ 的任意一组标准正交基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$,都有

$$(T\boldsymbol{e}_i, T\boldsymbol{e}_j) = (\boldsymbol{e}_i, \boldsymbol{e}_j) = \begin{cases} 1, & \text{当 } i = j \\ 0, & \text{当 } i \neq j \end{cases}$$

这里 $i, j = 1, 2, \cdots, n$,故 $T\boldsymbol{e}_1, T\boldsymbol{e}_2, \cdots, T\boldsymbol{e}_n$ 也为 $V$ 的一组标准正交基.

反之,当 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 是标准正交基时, $T\boldsymbol{e}_1, T\boldsymbol{e}_2, \cdots, T\boldsymbol{e}_n$ 也是标准正交基,则对 $V$ 中任意两个向量

$$\boldsymbol{\alpha} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n$$

$$\boldsymbol{\beta} = y_1\boldsymbol{e}_1 + y_2\boldsymbol{e}_2 + \cdots + y_n\boldsymbol{e}_n$$

便有

$$T\boldsymbol{\alpha} = x_1 T\boldsymbol{e}_1 + x_2 T\boldsymbol{e}_2 + \cdots + x_n T\boldsymbol{e}_n$$

$$T\boldsymbol{\beta} = y_1 T\boldsymbol{e}_1 + y_2 T\boldsymbol{e}_2 + \cdots + y_n T\boldsymbol{e}_n$$

因此

$$(T\boldsymbol{\alpha}, T\boldsymbol{\beta}) = x_1y_1 + x_2y_2 + \cdots + x_ny_n = (\boldsymbol{\alpha}, \boldsymbol{\beta})$$

从而 $T$ 是 $V$ 上的正交变换.

(iii)⇔(iv)

设 $T$ 在标准正交基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 下的矩阵为 $\boldsymbol{A}$,即有

$$T\boldsymbol{e}_i = \sum_{k=1}^{n} a_{ki}\boldsymbol{e}_k, \quad i = 1, 2, \cdots, n$$

若 $T\boldsymbol{e}_1, T\boldsymbol{e}_2, \cdots, T\boldsymbol{e}_n$ 也是标准正交基,则 $\boldsymbol{A}$ 可以看成由标准正交基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 到标准正交基的过渡矩阵,从而 $\boldsymbol{A}$ 是正交矩阵;反之,若 $\boldsymbol{A}$ 是正交矩阵,则

$$(T\boldsymbol{e}_i, T\boldsymbol{e}_j) = \left(\sum_{k=1}^{n} a_{ki}\boldsymbol{e}_k, \sum_{t=1}^{n} a_{tj}\boldsymbol{e}_t\right) = \sum_{k=1}^{n} a_{ki}a_{kj} = \begin{cases} 1, & \text{当 } i = j \\ 0, & \text{当 } i \neq j \end{cases}$$

最后一步是因为 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} = \boldsymbol{E}$ ($\boldsymbol{A}$ 为正交矩阵),这里 $i, j = 1, 2, \cdots, n$. 这说明 $T\boldsymbol{e}_1, T\boldsymbol{e}_2, \cdots, T\boldsymbol{e}_n$ 也为标准正交基.

这样,我们证明了(i)~(iv)的等价性. □

**例 2.5.1** 设 $\boldsymbol{G} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$,在 $\mathbf{R}^2$ 上定义旋转变换 $T$:对于任意 $\boldsymbol{x} \in \mathbf{R}^2$,定义 $T\boldsymbol{x} = \boldsymbol{G}\boldsymbol{x}$,则 $T$ 是在 $\mathbf{R}^2$ 上的正交变换.

**证** 由于正交矩阵 $G$ 是 $T$ 在标准正交基 $e_1=(1,0)^T$，$e_2=(0,1)^T$ 下的矩阵,所以 $T$ 是正交变换.

**例 2.5.2** [Householder(豪斯霍尔德)变换]在欧氏空间 $\mathbf{R}^n$ 中,对任一向量 $\boldsymbol{\alpha}\in\mathbf{R}^n$. 定义

$$H\boldsymbol{\alpha}=(\boldsymbol{E}-2\boldsymbol{\omega}\boldsymbol{\omega}^T)\boldsymbol{\alpha}$$

式中,$\boldsymbol{E}$ 为 $n$ 阶单位矩阵;$\boldsymbol{\omega}$ 是 $\mathbf{R}^n$ 中的单位列向量. 则 $H$ 是 $\mathbf{R}^n$ 的正交变换.

**证** 因为

$$H\boldsymbol{\alpha}=\boldsymbol{E}\boldsymbol{\alpha}-2\boldsymbol{\omega}\boldsymbol{\omega}^T\boldsymbol{\alpha}=\boldsymbol{\alpha}-2\boldsymbol{\omega}(\boldsymbol{\omega},\boldsymbol{\alpha})$$

所以

$$\begin{aligned}(H\boldsymbol{\alpha},H\boldsymbol{\alpha})&=(\boldsymbol{\alpha}-2\boldsymbol{\omega}(\boldsymbol{\omega},\boldsymbol{\alpha}),\boldsymbol{\alpha}-2\boldsymbol{\omega}(\boldsymbol{\omega},\boldsymbol{\alpha}))\\&=(\boldsymbol{\alpha},\boldsymbol{\alpha})-2(\boldsymbol{\alpha},(\boldsymbol{\omega},\boldsymbol{\alpha})\boldsymbol{\omega})-2((\boldsymbol{\omega},\boldsymbol{\alpha})\boldsymbol{\omega},\boldsymbol{\alpha})\\&\quad+4((\boldsymbol{\omega},\boldsymbol{\alpha})\boldsymbol{\omega},(\boldsymbol{\omega},\boldsymbol{\alpha})\boldsymbol{\omega})\\&=(\boldsymbol{\alpha},\boldsymbol{\alpha})\end{aligned}$$

从而 $H$ 是 $\mathbf{R}^n$ 的一个正交变换.

由定理 2.5.1 知,在标准正交基下,正交变换与正交矩阵对应. 由于正交矩阵的行列式等于+1 或−1,我们通常称行列式等于+1 的正交变换为旋转,或称为第一类的;行列式等于−1 的正交变换称为反射(镜像变换),或称为第二类的.

## 2.6 酉空间(复内积空间)简介

欧氏空间是专对实数域上线性空间而言. 虽然实内积空间理论在许多问题上扮演着重要角色,但是还不足以解决所有的问题. 复内积空间的引入可以弥补这一不足.

**定义 2.6.1** 设 $V$ 是复数域 $\mathbf{C}$ 上的线性空间,如果对 $V$ 中任意两个向量 $\boldsymbol{\alpha}$，$\boldsymbol{\beta}$ 都有一个复数(记为 $(\boldsymbol{\alpha},\boldsymbol{\beta})$)与它们对应,且满足下列条件,则复数 $(\boldsymbol{\alpha},\boldsymbol{\beta})$ 称 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$ 的内积:

(i) $(\boldsymbol{\alpha},\boldsymbol{\beta})=\overline{(\boldsymbol{\beta},\boldsymbol{\alpha})}$;

(ii) $(\lambda\boldsymbol{\alpha},\boldsymbol{\beta})=\lambda(\boldsymbol{\alpha},\boldsymbol{\beta})$，$\lambda\in\mathbf{C}$;

(iii) $(\boldsymbol{\alpha}+\boldsymbol{\beta},\boldsymbol{\gamma})=(\boldsymbol{\alpha},\boldsymbol{\gamma})+(\boldsymbol{\beta},\boldsymbol{\gamma})$，$\boldsymbol{\gamma}\in V$;

(iv) $(\boldsymbol{\alpha},\boldsymbol{\alpha})\geqslant 0$,当且仅当 $\boldsymbol{\alpha}=\mathbf{0}$ 时等号成立,

而 $V$ 称为复内积空间,或酉空间. 这里 $\overline{(\boldsymbol{\beta},\boldsymbol{\alpha})}$ 表示 $(\boldsymbol{\beta},\boldsymbol{\alpha})$ 的共轭复数.

**例 2.6.1** 在 $n$ 维线性空间 $\mathbf{C}^n$ 中,如果对 $\mathbf{C}^n$ 中任意二向量 $\boldsymbol{\alpha}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$, $\boldsymbol{\beta}=(y_1, y_2, \cdots, y_n)^{\mathrm{T}}$, 定义

$$(\boldsymbol{\alpha}, \boldsymbol{\beta})=x_1\bar{y}_1+x_2\bar{y}_2+\cdots+x_n\bar{y}_n$$

则容易验证$(\boldsymbol{\alpha}, \boldsymbol{\beta})$满足内积定义的各项条件,从而 $\mathbf{C}^n$ 构成一酉空间.

注意,内积定义中的条件(i)保证$(\boldsymbol{\alpha}, \boldsymbol{\alpha})$是实数. 又若条件(i)不作这样规定,而仍用欧氏空间定义中条件(i)的规定,就可能出现 $\boldsymbol{\alpha}\neq\mathbf{0}$ 但$(\boldsymbol{\alpha}, \boldsymbol{\alpha})=0$ 的情形. 例如,在 $\mathbf{C}^3$ 中,取 $\boldsymbol{\alpha}=(3, 4, 5\mathrm{i})\neq\mathbf{0}$,又定义$(\boldsymbol{\alpha}, \boldsymbol{\beta})=x_1y_1+x_2y_2+x_3y_3$,则有$(\boldsymbol{\alpha}, \boldsymbol{\alpha})=0$; 但按例 2.6.1 中定义 $(\boldsymbol{\alpha}, \boldsymbol{\alpha})\neq 0$.

酉空间 $\mathbf{C}^n$ 中内积的定义可以简记为

$$(\boldsymbol{\alpha}, \boldsymbol{\beta})=\boldsymbol{\alpha}\boldsymbol{\beta}^{\mathrm{H}}$$

这里 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 均为 $n$ 维行向量;$\boldsymbol{\beta}^{\mathrm{H}}$ 表示 $\boldsymbol{\beta}$ 的共轭转置. 当 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 为列向量时,则有 $(\boldsymbol{\alpha}, \boldsymbol{\beta})=\boldsymbol{\beta}^{\mathrm{H}}\boldsymbol{\alpha}$, 这种表示法以后经常会用到.

在例 2.6.1 中,若定义 $(\boldsymbol{\alpha}, \boldsymbol{\beta})=\bar{x}_1y_1+\bar{x}_2y_2+\cdots+\bar{x}_ny_n$, 则内积定义的条件(ii)便不满足.

由于酉空间的讨论与欧氏空间的讨论相似,因此本节只简单地列出重要的结论,而不详细论证.

首先由内积的定义可以得到:

(i) $(\boldsymbol{\alpha}, \lambda\boldsymbol{\beta})=\bar{\lambda}(\boldsymbol{\alpha}, \boldsymbol{\beta})$;

(ii) $(\boldsymbol{\alpha}, \boldsymbol{\beta}+\boldsymbol{\gamma})=(\boldsymbol{\alpha}, \boldsymbol{\beta})+(\boldsymbol{\alpha}, \boldsymbol{\gamma})$;

(iii) $(\boldsymbol{\alpha}, \mathbf{0})=(\mathbf{0}, \boldsymbol{\beta})=0$.

和在欧氏空间中一样,因为 $(\boldsymbol{\alpha}, \boldsymbol{\alpha})\geqslant 0$, 故可以定义向量的长度:

(iv) $\|\boldsymbol{\alpha}\|=\sqrt{(\boldsymbol{\alpha}, \boldsymbol{\alpha})}$;

(v) 柯西-施瓦茨不等式成立. 即对任意 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 有 $|(\boldsymbol{\alpha}, \boldsymbol{\beta})|\leqslant\|\boldsymbol{\alpha}\|\|\boldsymbol{\beta}\|$, 当且仅当 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$ 线性相关时,等号成立.

注意: 酉空间中的内积$(\boldsymbol{\alpha}, \boldsymbol{\beta})$一般是复数,故向量之间不再定义夹角,但我们仍引入:

(vi) 向量 $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$,当 $(\boldsymbol{\alpha}, \boldsymbol{\beta})=0$ 时称为正交或相互垂直;

在 $n$ 维酉空间中,同样可以定义正交基和标准正交基,并且关于标准正交基也有下述的重要性质:

(vii) 任意一组线性无关的向量可以用施密特过程正交化,并扩充成为一组标准正交基;

(viii) $n$ 维酉空间中,在标准正交基 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 下的任二向量

$$\boldsymbol{\alpha} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n$$

$$\boldsymbol{\beta} = y_1\boldsymbol{e}_1 + y_2\boldsymbol{e}_2 + \cdots + y_n\boldsymbol{e}_n$$

的内积

$$(\boldsymbol{\alpha}, \boldsymbol{\beta}) = x_1\bar{y}_1 + x_2\bar{y}_2 + \cdots + x_n\bar{y}_n$$

(ix) 对 $n$ 阶复矩阵 $\boldsymbol{A}$,若 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{E}$,则称 $\boldsymbol{A}$ 为酉矩阵,其中 $\boldsymbol{A}^{\mathrm{H}}$ 表示 $\boldsymbol{A}$ 的共轭转置.

两组标准正交基的过渡矩阵是酉矩阵.

**定义 2.6.2** 若 $T$ 为酉空间 $V$ 的线性变换,且对任何 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in V$,都有

$$(T\boldsymbol{\alpha}, T\boldsymbol{\beta}) = (\boldsymbol{\alpha}, \boldsymbol{\beta})$$

则称 $T$ 为 $V$ 的酉变换.

同欧氏空间情形一样,我们可以证明下述定理.

**定理 2.6.1** 设 $T$ 是 $n$ 维酉空间的 $V$ 的线性变换,则下列命题是等价的:

(i) $T$ 是酉变换;

(ii) $T$ 保持向量的长度不变,即 $\boldsymbol{\alpha} \in V$ 时,有

$$\| T\boldsymbol{\alpha} \| = \| \boldsymbol{\alpha} \|$$

(iii) 若 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 为 $V$ 的标准正交基,则 $T\boldsymbol{e}_1, T\boldsymbol{e}_2, \cdots, T\boldsymbol{e}_n$ 也是 $V$ 的标准正交基;

(iv) $T$ 在任一标准正交基下的矩阵是酉矩阵.

同欧氏空间情况类似,酉空间也有同构的概念,且所有 $n$ 维酉空间都与酉空间 $\mathbf{C}^n$ 同构. 因此,$n$ 维欧氏空间 $\mathbf{R}^n$ 和 $n$ 维酉空间 $\mathbf{C}^n$ 就成了 $n$ 维实内积空间和 $n$ 维复内积空间的代表,两者在应用中都很重要.

## 2.7 正规变换与正规矩阵

正规变换所对应的矩阵是一种具有特别性质及广泛应用的矩阵.

**定义 2.7.1** 设 $V$ 是 $n$ 维内积空间,$T$: $V \to V$ 是线性变换,如果存在 $V$ 的标准正交基$\{\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n\}$使得

$$(T(\boldsymbol{\varepsilon}_1), T(\boldsymbol{\varepsilon}_2), \cdots, T(\boldsymbol{\varepsilon}_n)) = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n)\begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix}$$

则称 $T$ 是 $V$ 上的正规变换.

若 $T$ 是正规变换,$T$ 在标准正交基$\{\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n\}$下的矩阵是 $\boldsymbol{A}$. 又假定$\{\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, \cdots, \tilde{\boldsymbol{\varepsilon}}_n\}$是 $V$ 的另一标准正交基,且

$$(\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, \cdots, \tilde{\boldsymbol{\varepsilon}}_n) = (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n)\boldsymbol{U} \tag{2.7.1}$$

则易知过渡矩阵 $\boldsymbol{U}$ 为酉矩阵. 此时,若设 $T$ 在$\{\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, \cdots, \tilde{\boldsymbol{\varepsilon}}_n\}$下的表示矩阵是 $\boldsymbol{B}$,即

$$(T(\tilde{\boldsymbol{\varepsilon}}_1), T(\tilde{\boldsymbol{\varepsilon}}_2), \cdots, T(\tilde{\boldsymbol{\varepsilon}}_n)) = (\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, \cdots, \tilde{\boldsymbol{\varepsilon}}_n)\boldsymbol{B} \tag{2.7.2}$$

则由式(2.7.1)知

$$\begin{aligned}(T(\tilde{\boldsymbol{\varepsilon}}_1), T(\tilde{\boldsymbol{\varepsilon}}_2), \cdots, T(\tilde{\boldsymbol{\varepsilon}}_n)) &= (T(\boldsymbol{\varepsilon}_1), T(\boldsymbol{\varepsilon}_2), \cdots, T(\boldsymbol{\varepsilon}_n))\boldsymbol{U} \\ &= (\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \cdots, \boldsymbol{\varepsilon}_n)\boldsymbol{AU} \\ &= (\tilde{\boldsymbol{\varepsilon}}_1, \tilde{\boldsymbol{\varepsilon}}_2, \cdots, \tilde{\boldsymbol{\varepsilon}}_n)\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AU}\end{aligned}$$

从而有

$$\boldsymbol{B} = \boldsymbol{U}^{\mathrm{H}}\boldsymbol{AU}$$

我们称 $\boldsymbol{A}$ 与 $\boldsymbol{B}$ 是酉相似的.

上述结果表明,正规变换在不同标准正交基下的表示矩阵是酉相似矩阵. 而 $T$ 在任一标准正交基下的表示矩阵必定酉相似于对角阵.

**定义 2.7.2** 设矩阵 $\boldsymbol{A}$ 为一正规变换 $T$ 在一标准正交基下的矩阵,则称 $\boldsymbol{A}$ 为正规矩阵.

一个给定的方阵是否为正规矩阵有更为简便的判别方法. 为此我们先介绍一个定理.

**定理 2.7.1** [Schur(舒尔)引理]任意 $n$ 阶方阵 $\boldsymbol{A}$ 必酉相似于一个上三角阵 $\boldsymbol{T}$. 即存在酉矩阵 $\boldsymbol{U}$,使得

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{AU} = \boldsymbol{T}$$

**证** 设 $\lambda_1$ 是 $\boldsymbol{A}$ 的一个特征值,$\boldsymbol{\gamma}_1$ 是对应于 $\lambda_1$ 的单位特征向量. 将 $\boldsymbol{\gamma}_1$ 扩充成 $\mathbf{C}^n$ 的标准正交基:$\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n$. 取 $\boldsymbol{U}_1 = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n)$,则

$$\boldsymbol{AU}_1 = \boldsymbol{U}_1\left(\begin{array}{c|ccc}\lambda_1 & * & \cdots & * \\ \hline \boldsymbol{0} & & \boldsymbol{A}_{n-1} & \end{array}\right) \quad 或 \quad \boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{AU}_1 = \left(\begin{array}{c|ccc}\lambda_1 & * & \cdots & * \\ \hline \boldsymbol{0} & & \boldsymbol{A}_{n-1} & \end{array}\right)$$

对矩阵 $\boldsymbol{A}_{n-1}$ 进行类似的过程. 设 $\lambda_2$ 是 $\boldsymbol{A}_{n-1}$ 的一个特征值,$\boldsymbol{\eta}_2$ 是对应于 $\lambda_2$ 的单位特征向量. 将 $\boldsymbol{\eta}_2$ 扩充成 $\mathbf{C}^{n-1}$ 的标准基 $\boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \cdots, \boldsymbol{\eta}_n$. 令 $\tilde{\boldsymbol{U}}_2 = (\boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \cdots, \boldsymbol{\eta}_n)$,则

$$\widetilde{\boldsymbol{U}}_2^{\mathrm{H}}\boldsymbol{A}_{n-1}\widetilde{\boldsymbol{U}}_2=\left(\begin{array}{c:ccc}\lambda_2 & * & \cdots & * \\ \hdashline \boldsymbol{0} & & \boldsymbol{A}_{n-2} & \end{array}\right)$$

再取

$$\boldsymbol{U}_2=\begin{pmatrix}1 & \boldsymbol{0}^{\mathrm{T}} \\ \boldsymbol{0} & \widetilde{\boldsymbol{U}}_2\end{pmatrix}$$

则有

$$\boldsymbol{U}_2^{\mathrm{H}}(\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}_1)\boldsymbol{U}_2=\left(\begin{array}{cc:ccc}\lambda_1 & * & * & \cdots & * \\ 0 & \lambda_2 & * & \cdots & * \\ \hdashline \boldsymbol{0} & \boldsymbol{0} & & \boldsymbol{A}_{n-2} & \end{array}\right)$$

继续上述过程,直到找到 $n-1$ 个酉矩阵 $\boldsymbol{U}_1, \boldsymbol{U}_2, \cdots, \boldsymbol{U}_{n-1}$ 使得

$$\boldsymbol{U}_{n-1}^{\mathrm{H}}\boldsymbol{U}_{n-2}^{\mathrm{H}}\cdots\boldsymbol{U}_1^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}_1\cdots\boldsymbol{U}_{n-1}=\begin{pmatrix}\lambda_1 & & & & * \\ & \lambda_2 & & & \\ & & \ddots & & \\ & & & \lambda_{n-1} & \\ & & & & \lambda_n\end{pmatrix}=\boldsymbol{T}$$

此时,只要取 $\boldsymbol{U}=\boldsymbol{U}_1\boldsymbol{U}_1\cdots\boldsymbol{U}_{n-1}$,则 $\boldsymbol{U}$ 就是 $n$ 阶酉矩阵,且使

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{U}=\boldsymbol{T}$$

$\boldsymbol{T}$ 的主对角元全为 $\boldsymbol{A}$ 的特征值. □

Schur 引理是 1909 年给出的,该结论在证明正规矩阵的一个美妙特征时起着至关重要的作用.下面的定理常被用做正规矩阵的等价定义.

**定理 2.7.2** $n\times n$ 矩阵 $\boldsymbol{A}$ 是正规的,当且仅当

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$$

**证** 设 $\boldsymbol{A}$ 为正规矩阵,则 $\boldsymbol{A}$ 必酉相似于对角阵.即存在 $n$ 阶酉矩阵 $\boldsymbol{U}$ 及对角阵 $\boldsymbol{D}$,使得 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}$. 于是

$$\begin{aligned}\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}&=(\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}})(\boldsymbol{U}\bar{\boldsymbol{D}}\boldsymbol{U}^{\mathrm{H}})=\boldsymbol{U}\boldsymbol{D}\bar{\boldsymbol{D}}\boldsymbol{U}^{\mathrm{H}}=\boldsymbol{U}\bar{\boldsymbol{D}}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}\\&=(\boldsymbol{U}\bar{\boldsymbol{D}}\boldsymbol{U}^{\mathrm{H}})(\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}})=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\end{aligned}$$

反之,假定 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$. 应用定理 2.7.1,存在酉矩阵 $\boldsymbol{U}$,及上三角阵 $\boldsymbol{T}=(t_{ij})_n$ 使得 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{T}\boldsymbol{U}^{\mathrm{H}}$.

容易看出,$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 当且仅当 $\boldsymbol{T}\boldsymbol{T}^{\mathrm{H}}=\boldsymbol{T}^{\mathrm{H}}\boldsymbol{T}$. 比较两边矩阵在第 $i$ 行,第 $i$ 列

位置的元素，并注意到 $t_{ij}=0$，当 $i>j$ 时，则

$$\begin{aligned}&|t_{ii}|^2+|t_{i,i+1}|^2+\cdots+|t_{in}|^2\\=&|t_{1i}|^2+|t_{2i}|^2+\cdots+|t_{ii}|^2,\quad i=1,2,\cdots,n\end{aligned}$$

当 $i=1$ 时，即有

$$|t_{11}|^2+|t_{12}|^2+\cdots+|t_{1n}|^2=|t_{11}|^2 \tag{2.7.3}$$

由式(2.7.3)即知 $t_{1j}=0,\ j=2,3,\cdots,n.$

当 $i=2$ 时，即有

$$|t_{22}|^2+|t_{23}|^2+\cdots+|t_{2n}|^2=|t_{12}|^2+|t_{22}|^2 \tag{2.7.4}$$

因 $t_{12}=0$，故由式(2.7.4)知 $t_{2j}=0,\ j=3,4,\cdots,n.$

继续这一过程可以求得 $t_{ij}=0,\ i<j$. 所以 $\boldsymbol{T}$ 必定为对角阵，从而 $\boldsymbol{A}$ 是正规矩阵. □

利用定理 2.7.2 来判断方阵是否为正规矩阵是十分简便的. 由该判别法容易看出：实对称矩阵 ($\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}$)、实反对称矩阵($\boldsymbol{A}^{\mathrm{T}}=-\boldsymbol{A}$)、正交矩阵($\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{A}^{-1}$)、酉矩阵 ($\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}^{-1}$)、Hermite(埃尔米特)矩阵 ($\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$)、反 Hermite 矩阵 ($\boldsymbol{A}^{\mathrm{H}}=-\boldsymbol{A}$) 都是正规矩阵. 但应注意，也有正规矩阵不是上面的任一种类型. 如

$$\boldsymbol{A}=\begin{pmatrix}1&-1\\1&1\end{pmatrix}$$

在所有的正规矩阵中，Hermite 矩阵与反 Hermite 矩阵是结构与性质比较奇特的矩阵，而且这类矩阵我们会经常在实际应用中遇到.

以下我们给出关于 Hermite 矩阵及反 Hermite 矩阵的几个性质. 对 Hermite 矩阵的性质给出了证明，而关于反 Hermite 矩阵的对应性质请读者仿照完成.

**定理 2.7.3** Hermite 矩阵的特征值全为实数，反 Hermite 矩阵的特征值全为纯虚数.

**证** 设 $\boldsymbol{H}$ 是 Hermite 矩阵，则因 $\boldsymbol{H}$ 正规，故 $\boldsymbol{H}$ 必酉相似于对角阵. 即有酉矩阵 $\boldsymbol{U}$ 及对角阵 $\boldsymbol{D}$，使得

$$\boldsymbol{H}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}$$

显然，$\boldsymbol{D}$ 的主对角元是 $\boldsymbol{H}$ 的全部特征值.

由于 $\boldsymbol{H}^{\mathrm{H}}=\boldsymbol{H}$，从而

$$\boldsymbol{U}\bar{\boldsymbol{D}}\boldsymbol{U}^{\mathrm{H}}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}$$

由此得
$$\bar{\boldsymbol{D}}=\boldsymbol{D}$$

所以，$\boldsymbol{D}$ 的主对角元全为实数. □

**定理 2.7.4** 正规矩阵 $\boldsymbol{A}$ 是 Hermite(反 Hermite)矩阵，当且仅当 $\boldsymbol{A}$ 的特征值全为实数(纯虚数).

**证** 必要性已由定理 2.7.3 证明. 下证充分性.

假定 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 且 $\boldsymbol{A}$ 的特征值全为实数，则因

$$\boldsymbol{A}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}, \qquad \boldsymbol{A}^{\mathrm{H}}=\boldsymbol{U}\bar{\boldsymbol{D}}\boldsymbol{U}^{\mathrm{H}}=\boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}$$

所以

$$\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$$

即 $\boldsymbol{A}$ 是 Hermite 矩阵. □

由此可见，如果一个正规矩阵的特征值不全是实数，则可以断定该矩阵不是 Hermite 矩阵.

**定理 2.7.5** $n$ 阶方阵 $\boldsymbol{A}$ 是 Hermite 矩阵，当且仅当对任意的 $\boldsymbol{\alpha}, \boldsymbol{\beta}\in \mathbf{C}^n$，$(\boldsymbol{A\alpha}, \boldsymbol{\beta})=(\boldsymbol{\alpha}, \boldsymbol{A\beta})$.

**证** 设 $\boldsymbol{A}$ 是 Hermite 矩阵，$\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{A}$，则对任意的 $\boldsymbol{\alpha}, \boldsymbol{\beta}\in \mathbf{C}^n$，有

$$(\boldsymbol{A\alpha}, \boldsymbol{\beta})=\boldsymbol{\beta}^{\mathrm{H}}\boldsymbol{A\alpha}=\boldsymbol{\beta}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\alpha}=(\boldsymbol{\alpha}, \boldsymbol{A\beta})$$

反之，设上式成立，$\boldsymbol{A}=(a_{ij})_n$，$\boldsymbol{e}_i=(0, \cdots, 0, \overset{(i)}{1}, 0, \cdots, 0)^{\mathrm{T}}$，$i=1, 2, \cdots, n$，那么就有

$$(\boldsymbol{A}\boldsymbol{e}_j, \boldsymbol{e}_i)=\boldsymbol{e}_i^{\mathrm{H}}\boldsymbol{A}\boldsymbol{e}_j=a_{ij}; \qquad (\boldsymbol{e}_j, \boldsymbol{A}\boldsymbol{e}_i)=\boldsymbol{e}_i^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{e}_j=\bar{a}_{ji}$$

所以

$$a_{ij}=\bar{a}_{ji}, \quad i, j=1, 2, \cdots, n$$

即

$$\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$$ □

读者可以仿此给出关于反 Hermite 矩阵的对应结论.

对任一方阵 $\boldsymbol{A}$，如果我们定义

$$\boldsymbol{H}_1=\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}), \qquad \boldsymbol{H}_2=\frac{1}{2\mathrm{i}}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{H}})$$

则易知 $\boldsymbol{H}_1$，$\boldsymbol{H}_2$ 都是 Hermite 矩阵，并且

$$\boldsymbol{A}=\boldsymbol{H}_1+\mathrm{i}\boldsymbol{H}_2$$

上式称为 $\boldsymbol{A}$ 的笛卡儿(Cartesian)分解. 若 $\boldsymbol{A}$ 的阶数 $n=1$，则得到复数的代数表示式

$$z = a + \mathrm{i}b$$

在全体 Hermite 矩阵中，有些特殊的矩阵在以后会用到.

**定义 2.7.3** 若 Hermite 矩阵 $\boldsymbol{H}$ 的特征值全为正实数(非负实数)，则称 $\boldsymbol{H}$ 为正定(半正定)矩阵，记为 $\boldsymbol{H} > 0 (\geqslant 0)$.

**定理 2.7.6** 矩阵 $\boldsymbol{H}$ 是正定(或半正定)的充分必要条件是存在正定矩阵(或半正定矩阵)$\boldsymbol{H}_0$，使得

$$\boldsymbol{H} = \boldsymbol{H}_0^2$$

并且

$$\operatorname{rank} \boldsymbol{H}_0 = \operatorname{rank} \boldsymbol{H}$$

**证** 我们只对 $\boldsymbol{H} > 0$ 的情形给出证明. 设 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 为 $\boldsymbol{H}$ 的全部特征值，则由定义，$\lambda_i > 0$, $i = 1, 2, \cdots, n$，令 $D_0 = \mathrm{ding}(\sqrt{\lambda_1}, \sqrt{\lambda_2}, \cdots, \sqrt{\lambda_n})$, $\boldsymbol{D} = \boldsymbol{D}_0^2$. 则因 $\boldsymbol{H}$ 是正规的，由定理 2.7.5，存在一个酉矩阵 $\boldsymbol{U}$，使得

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}$$

于是当我们取 $\boldsymbol{H}_0 = \boldsymbol{U}\boldsymbol{D}_0\boldsymbol{U}^{\mathrm{H}}$ 时，就有

$$\boldsymbol{H}_0^2 = (\boldsymbol{U}\boldsymbol{D}_0\boldsymbol{U}^{\mathrm{H}})(\boldsymbol{U}\boldsymbol{D}_0\boldsymbol{U}^{\mathrm{H}}) = \boldsymbol{U}\boldsymbol{D}_0^2\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}} = \boldsymbol{H}$$

显然有 $\operatorname{rank} \boldsymbol{H}_0 = n = \operatorname{rank} \boldsymbol{H}$，而 $\boldsymbol{H}_0 > 0$.

反之，设有 $\boldsymbol{H}_0 > 0$ 使得 $\boldsymbol{H} = \boldsymbol{H}_0^2$，则 $\boldsymbol{H}_0$ 的特征值的平方是 $\boldsymbol{H}$ 的特征值. 但 $\boldsymbol{H}_0$ 的特征值全大于零，所以 $\boldsymbol{H}$ 的特征值全大于零，于是 $\boldsymbol{H} > 0$. □

**定理 2.7.7** 给定 $n$ 阶 Hermite 矩阵 $\boldsymbol{H}$，则 $\boldsymbol{H} \geqslant 0$(或 $\boldsymbol{H} > 0$) 当且仅当对任意非零 $n$ 维向量 $\boldsymbol{\alpha}$，有

$$(\boldsymbol{H}\boldsymbol{\alpha}, \boldsymbol{\alpha}) \geqslant 0 \quad (\text{或}(\boldsymbol{H}\boldsymbol{\alpha}, \boldsymbol{\alpha}) > 0)$$

**证** 设 $\boldsymbol{H}$ 是半正定的，则由定理 2.7.6，存在半正定矩阵 $H_0$，使得

$$\boldsymbol{H} = \boldsymbol{H}_0^2$$

于是根据定理 2.7.5，有

$$(\boldsymbol{H}\boldsymbol{\alpha}, \boldsymbol{\alpha}) = (\boldsymbol{H}_0^2\boldsymbol{\alpha}, \boldsymbol{\alpha}) = (\boldsymbol{H}_0\boldsymbol{\alpha}, H_0\boldsymbol{\alpha}) \geqslant 0$$

反之，若对任意 $\boldsymbol{\alpha} \neq \boldsymbol{0}$, $(\boldsymbol{H}\boldsymbol{\alpha}, \boldsymbol{\alpha}) \geqslant 0$. 假定 $\boldsymbol{u}_1, \boldsymbol{u}_2, \cdots, \boldsymbol{u}_n$ 是单位正交且分别对应于 $\boldsymbol{H}$ 的特征值 $\lambda_1, \lambda_2, \cdots, \lambda_n$，则因

$$0 \leqslant (\boldsymbol{H}\boldsymbol{u}_i, \boldsymbol{u}_i) = (\lambda_i \boldsymbol{u}_i, \boldsymbol{u}_i) = \lambda_i(\boldsymbol{u}_i, \boldsymbol{u}_i) = \lambda_i, \quad i = 1, 2, \cdots, n$$

所以由定义 2.7.3 知

$$\boldsymbol{H} \geqslant 0$$

□

对半正定(或正定)矩阵 $\boldsymbol{H}$,我们把满足定理 2.7.6 中的条件 $\boldsymbol{H}=\boldsymbol{H}_0^2$ 的矩阵 $\boldsymbol{H}_0$ 称为 $\boldsymbol{H}$ 的方根,并记为

$$\boldsymbol{H}_0=\boldsymbol{H}^{\frac{1}{2}}$$

现在考察任一 $m\times n$ 矩阵 $\boldsymbol{A}$,则 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 分别为 $n$ 阶与 $m$ 阶 Hermite 矩阵. 并且对任意非零 $n$ 维向量 $\boldsymbol{\alpha}$,因

$$(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{\alpha},\ \boldsymbol{\alpha})=\boldsymbol{\alpha}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{\alpha}=(\boldsymbol{A}\boldsymbol{\alpha})^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{\alpha})\geqslant 0$$

所以,$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 必为半正定(或正定)矩阵. 同理,$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 也是半正定(或正定)矩阵.

**定义 2.7.4** 设 $\boldsymbol{A}$ 为 $m\times n$ 矩阵,则 $n$ 阶矩阵 $(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{\frac{1}{2}}$ 的特征值称为 $\boldsymbol{A}$ 的奇异值.

尽管 $(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{\frac{1}{2}}$ 与 $(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{\frac{1}{2}}$ 是不同的,但我们可以证明它们的非零特征值相同.

**定理 2.7.8** $(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{\frac{1}{2}}$ 与 $(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{\frac{1}{2}}$ 的非零特征值相同.

**证** 因为 $(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{\frac{1}{2}}$ 与 $(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{\frac{1}{2}}$ 的特征值必分别是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的特征值的平方,所以,我们只要证明 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 有相同的非零特征值.

设 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{\alpha}=\lambda\boldsymbol{\alpha},\ \lambda\neq 0,\ \boldsymbol{\alpha}\neq\boldsymbol{0}$,则必有 $\boldsymbol{\beta}=\boldsymbol{A}\boldsymbol{\alpha}\neq\boldsymbol{0}$. 此时,因

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{\beta}=\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{\alpha})=\lambda\boldsymbol{A}\boldsymbol{\alpha}=\lambda\boldsymbol{\beta}$$

故 $\lambda$ 是 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的特征值. 同理可证 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 的非零特征值也是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的非零特征值. □

注意,$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 与 $\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$ 未必有相同重数的零特征值. 如

$$\boldsymbol{A}=\begin{pmatrix}1\\0\end{pmatrix}$$

时

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=(1),\qquad \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}=\begin{pmatrix}1&0\\0&0\end{pmatrix}$$

**例 2.7.1** 设矩阵

$$\boldsymbol{A}=\begin{bmatrix}1&0\\-1&1\\-1&1\end{bmatrix}$$

试求 $\boldsymbol{A}$ 的奇异值.

**解** 因为

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}=\begin{pmatrix}3&-2\\-2&2\end{pmatrix}$$

的特征值是 $\frac{5\pm\sqrt{17}}{2}$，所以，$\boldsymbol{A}$ 的奇异值是

$$\sqrt{\frac{5+\sqrt{17}}{2}},\qquad \sqrt{\frac{5-\sqrt{17}}{2}}$$

## 习题 2

1. 设 $\boldsymbol{\alpha}=(x_1,x_2)^{\mathrm{T}}$，$\boldsymbol{\beta}=(y_1,y_2)^{\mathrm{T}}\in\mathbf{R}^2$，试问 $\mathbf{R}^2$ 对以下定义的内积是否构成欧氏空间：

(1) $(\boldsymbol{\alpha},\boldsymbol{\beta})=x_1y_1+x_2y_2+1$；

(2) $(\boldsymbol{\alpha},\boldsymbol{\beta})=x_1y_1-x_2y_2$；

(3) $(\boldsymbol{\alpha},\boldsymbol{\beta})=3x_1y_1+5x_2y_2$；

(4) $(\boldsymbol{\alpha},\boldsymbol{\beta})=x_1y_2+x_2y_1$.

2. 对 $\mathbf{R}^3$ 中任意向量

$$\boldsymbol{\alpha}=(x_1,x_2,x_3)^{\mathrm{T}},\qquad \boldsymbol{\beta}=(y_1,y_2,y_3)^{\mathrm{T}}$$

定义

$$(\boldsymbol{\alpha},\boldsymbol{\beta})=x_1y_1+2x_2y_2+3x_3y_3$$

证明：这是 $\mathbf{R}^3$ 上的内积；并求 $\mathbf{R}^3$ 的一组标准正交基.

3. 设 $\boldsymbol{A}$ 是一个 $n$ 阶实对称矩阵，而

$$\boldsymbol{\alpha}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}},\qquad \boldsymbol{\beta}=(y_1,y_2,\cdots,y_n)^{\mathrm{T}}\in\mathbf{R}^n$$

定义 $(\boldsymbol{\alpha},\boldsymbol{\beta})=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\beta}$. 证明 $\mathbf{R}^n$ 对所定义的 $(\boldsymbol{\alpha},\boldsymbol{\beta})$ 构成欧氏空间的充要条件是 $\boldsymbol{A}$ 为正定矩阵.

4. 设 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\cdots,\boldsymbol{\alpha}_n$ 是 $n$ 维欧氏空间 $V$ 的一组基，证明：

(1) 如果 $\boldsymbol{\gamma}\in V$ 使 $(\boldsymbol{\gamma},\boldsymbol{\alpha}_i)=0$，$i=1,2,\cdots,n$，那么 $\boldsymbol{\gamma}=\mathbf{0}$；

(2) 如果 $\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2\in V$ 使对任一 $\boldsymbol{\alpha}\in V$ 有

$$(\boldsymbol{\gamma}_1,\boldsymbol{\alpha})=(\boldsymbol{\gamma}_2,\boldsymbol{\alpha})$$

则

$$\boldsymbol{\gamma}_1=\boldsymbol{\gamma}_2$$

5. 求齐次线性方程组

$$\begin{cases}2x_1+x_2-x_3+x_4-3x_5=0\\ x_1+x_2-x_3+x_5=0\end{cases}$$

的解空间的一组标准正交基

6. 设 $V$ 是 $n$ 维欧氏空间，$\boldsymbol{\alpha}$ 为 $V$ 中一个取定的非零向量，证明：

(1) $V_1 = \{\boldsymbol{\beta} \in V \mid (\boldsymbol{\alpha}, \boldsymbol{\beta}) = 0\}$ 是 $V$ 的子空间；

(2) $\dim V_1 = n-1$.

7. 设 $\boldsymbol{\alpha} = (1, -2, 1)^{\mathrm{T}} \in \mathbf{R}^3$，$W = \mathrm{Span}\{\boldsymbol{\alpha}\}$，试求 $W^{\perp}$.

8. 证明关于正交补子空间的下列性质：

(1) $(V_0^{\perp})^{\perp} = V_0^{\perp}$；

(2) 若 $V_1 \subset V_2$，则 $V_1^{\perp} \supset V_2^{\perp}$；

(3) $(V_1 + V_2)^{\perp} = V_1^{\perp} \cap V_2^{\perp}$；

(4) $(V_1 \cap V_2)^{\perp} = V_1^{\perp} + V_2^{\perp}$.

9. 证明弗雷德霍姆(Fredholm)定理：线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}(\boldsymbol{A} \in \mathbf{C}^{m\times n})$ 有解的充分必要条件是向量 $\boldsymbol{b} \in \mathbf{C}^m$ 与齐次线性方程组 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{y} = \boldsymbol{0}$ 的解空间正交.

10. 设 $T$ 是欧氏空间 $V$ 上的正交变换，$V$ 的两个子空间分别为

$$V_1 = \{\boldsymbol{x} \mid T\boldsymbol{x} = \boldsymbol{x},\ \boldsymbol{x} \in V\}$$

$$V_2 = \{\boldsymbol{y} \mid \boldsymbol{y} = \boldsymbol{x} - T\boldsymbol{x},\ \boldsymbol{x} \in V\}$$

证明：$V_1 = V_2^{\perp}$.

11. 证明下面命题等价：

(1) $\boldsymbol{A}$ 为酉矩阵，即 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{E}$；

(2) $\boldsymbol{A}^{-1} = \boldsymbol{A}^{\mathrm{H}}$；

(3) $\boldsymbol{A}$ 的列向量组是 $\mathbf{C}^n$ 的标准正交向量组；

(4) $\boldsymbol{A}$ 的行向量组是 $\mathbf{C}^n$ 的标准正交向量组.

12. 已知 $\boldsymbol{A} = \begin{pmatrix} -1 & \mathrm{i} & 0 \\ -\mathrm{i} & 0 & -\mathrm{i} \\ 0 & \mathrm{i} & -1 \end{pmatrix}$，试问 $\boldsymbol{A}$ 是否为正规矩阵，若是，试求酉矩阵 $\boldsymbol{U}$，使 $\boldsymbol{U}^{-1}\boldsymbol{A}\boldsymbol{U}$ 为对角矩阵.

13. 设 $\boldsymbol{x}_1, \boldsymbol{x}_2$ 是方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的两个最小二乘解，证明：$\boldsymbol{A}\boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{x}_2$.

# 第3章 矩阵的标准形

在线性代数中我们已经知道，不是任何一个 $n$ 阶矩阵都与对角矩阵相似的. 不与对角矩阵相似的矩阵能与何种具有“标准”形状的矩阵相似呢？这个问题就是本章要讨论的矩阵在相似下的 Jordan(若尔当)标准形问题.

矩阵的标准形问题不仅在矩阵理论和矩阵计算中有着重要地位，而且在力学、控制理论、系统分析等领域也有着广泛的应用.

因课时所限，有些问题仅叙述结论而略去证明.

## 3.1 Jordan 标准形

虽然一个 $n$ 阶矩阵不一定相似于对角阵，但这个 $n$ 阶矩阵能相似于一个形式上比对角阵稍复杂的 Jordan 标准形 $\boldsymbol{J}$. 由于 Jordan 标准形的独特结构揭示了两个矩阵相似的本质，故在数值计算和理论推导中经常采用.

**定义 3.1.1** 形如

$$\boldsymbol{J}_i=\begin{pmatrix}\lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & & \ddots & \\ & & \ddots & & 1\\ & & & & \lambda_i\end{pmatrix}_{S_i\times S_i}$$

的方阵称为 $S_i$ 阶 Jordan 块，其中 $\lambda_i$ 可以是实数，也可以是复数.

例如

$$\begin{pmatrix}2 & 1\\ 0 & 2\end{pmatrix},\quad \begin{pmatrix}1+\mathrm{i} & 1 & 0\\ 0 & 1+\mathrm{i} & 1\\ 0 & 0 & 1+\mathrm{i}\end{pmatrix},\quad \begin{pmatrix}0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 0\end{pmatrix}$$

都是 Jordan 块. 特别地，一阶方阵是一阶 Jordan 块.

**定义 3.1.2** 由若干个 Jordan 块组成的分块对角阵为

$$\boldsymbol{J}=\begin{pmatrix}\boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_t\end{pmatrix}$$

其中 $\boldsymbol{J}_i(i=1,\ 2,\ \cdots,\ t)$ 为 $S_i$ 阶 Jordan 块，当 $\sum_{i=1}^{t}S_i=n$ 时，称 $\boldsymbol{J}$ 为 $n$ 阶 Jordan 标准形.

例如

$$\boldsymbol{J}=\left(\begin{array}{c|cc|ccc}1 & & & & & \\ \hline & 1+i & 1 & & & \\ & 0 & 1+i & & & \\ \hline & & & 0 & 1 & 0 \\ & & & 0 & 0 & 1 \\ & & & 0 & 0 & 0\end{array}\right)$$

是 6 阶 Jordan 标准形.

特别地，对角阵

$$\boldsymbol{A}=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}$$

也是 Jordan 标准形，其中每个 Jordan 块都是一阶的.

下面讨论任何一个矩阵 $\boldsymbol{A}$ 与 Jordan 标准形 $\boldsymbol{J}$ 相似的条件，以及如何将矩阵 $\boldsymbol{A}$ 化为 Jordan 标准形 $\boldsymbol{J}$.

设 $\boldsymbol{A}=(a_{ij})\in\mathbf{C}^{n\times n}$，$\lambda\boldsymbol{E}-\boldsymbol{A}$ 是 $\boldsymbol{A}$ 的特征矩阵，记为 $\boldsymbol{A}(\lambda)$.

**定义 3.1.3** $\boldsymbol{A}(\lambda)$ 中所有非零的 $k$ 阶子式的首项(最高次项)系数为 1 的最大公因式 $D_k(\lambda)$ 称为 $\boldsymbol{A}(\lambda)$ 的一个 $k$ 级行列式因子，$k=1,\ 2,\ \cdots,\ n$.

由定义可知，$D_n(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}|$. 又因为 $D_{k-1}(\lambda)$ 能整除每个 $k-1$ 级子式，从而可以整除每个 $k$ 级子式，因此 $D_{k-1}(\lambda)$ 能整除 $D_k(\lambda)$，并记为 $D_{k-1}(\lambda)\mid D_k(\lambda)$ $(k=2,\ 3,\ \cdots,\ n)$.

**定义 3.1.4** 下列 $n$ 个多项式：

$$d_1(\lambda)=D_1(\lambda),\ d_2(\lambda)=\frac{D_2(\lambda)}{D_1(\lambda)},\ d_3(\lambda)=\frac{D_3(\lambda)}{D_2(\lambda)},\ \cdots,\ d_n(\lambda)=\frac{D_n(\lambda)}{D_{n-1}(\lambda)}$$

称为 $\boldsymbol{A}(\lambda)$的不变因式.

**定义 3.1.5**　将 $\boldsymbol{A}(\lambda)$的每个次数大于零的不变因式分解为互不相同的一次因式的方幂的乘积(因在复数域内讨论,这样分解可行),所有这些一次因式方幂(相同的必须按出现次数计算),称为 $\boldsymbol{A}(\lambda)$的初级因子.

在这里,由于 $\boldsymbol{A}(\lambda)=\lambda\boldsymbol{E}-\boldsymbol{A}$ 完全由矩阵 $\boldsymbol{A}$ 确定,所以这里 $\boldsymbol{A}(\lambda)$的不变因子及初级因子,也称为矩阵 $\boldsymbol{A}$ 的不变因子及初级因子.

**例 3.1.1**　求矩阵

$$\boldsymbol{A}=\begin{pmatrix}1 & & \\ & -2 & \\ & & -2\end{pmatrix}$$

的不变因子及初级因子.

**解**　$\boldsymbol{A}$ 的特征矩阵为

$$\boldsymbol{A}(\lambda)=\lambda\boldsymbol{E}-\boldsymbol{A}=\begin{pmatrix}\lambda-1 & & \\ & \lambda+2 & \\ & & \lambda+2\end{pmatrix}$$

由此 $\boldsymbol{A}$ 的行列式因子为

$$D_1(\lambda)=1,\quad D_2(\lambda)=\lambda+2,\quad D_3(\lambda)=(\lambda+2)^2(\lambda-1)$$

$\boldsymbol{A}$ 的不变因子为

$$d_1(\lambda)=1,\quad d_2(\lambda)=\lambda+2,\quad d_3(\lambda)=(\lambda+2)(\lambda-1)$$

$\boldsymbol{A}$ 的全部初级因子为

$$\lambda+2,\quad \lambda+2,\quad \lambda-1$$

**例 3.1.2**　设

$$\boldsymbol{A}=\begin{pmatrix}-1 & -2 & 6\\ -1 & 0 & 3\\ -1 & -1 & 4\end{pmatrix}$$

试求 $\boldsymbol{A}$ 的初级因子.

**解**　因为

$$A(\lambda)=|\lambda E-A|=\begin{vmatrix}\lambda+1 & -2 & -6\\ 1 & \lambda & -3\\ 1 & 1 & \lambda-4\end{vmatrix}$$

可以求得

$$D_1(\lambda)=1,\quad D_2(\lambda)=\lambda-1,\quad D_3(\lambda)=(\lambda-1)^3$$

所以不变因子为

$$d_1(\lambda)=1,\quad d_2(\lambda)=\lambda-1,\quad d_3(\lambda)=(\lambda-1)^2$$

则 $A$ 的初级因子为

$$\lambda-1,\quad (\lambda-1)^2$$

**例 3.1.3** 设

$$A=\begin{pmatrix}a & 1 & 0 & \cdots & 0 & 0\\ 0 & a & 1 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & a & 1\\ 0 & 0 & 0 & \cdots & 0 & a\end{pmatrix}_k$$

试求 $A$ 的初级因子.

**解** 因为

$$A(\lambda)=\lambda E-A=\begin{pmatrix}\lambda-a & -1 & 0 & \cdots & 0 & 0\\ 0 & \lambda-a & -1 & \cdots & 0 & 0\\ \vdots & \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & 0 & \cdots & \lambda-a & -1\\ 0 & 0 & 0 & \cdots & 0 & \lambda-a\end{pmatrix}_k$$

所以

$$D_k(\lambda)=|\lambda E-A|=(\lambda-a)^k$$

又由

$$\begin{vmatrix}-1 & 0 & \cdots & 0 & 0\\ \lambda-a & -1 & \cdots & 0 & 0\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & \lambda-a & -1\end{vmatrix}_{k-1}=(-1)^{k-1}\neq 0$$

有 $D_{k-1}(\lambda)=1$，从而

$$D_{k-2}(\lambda) = \cdots = D_1(\lambda) = 1$$

于是不变因式为

$$d_1(\lambda) = \cdots = d_{k-1}(\lambda) = 1, \quad d_k(\lambda) = (\lambda - a)^k$$

因此初级因子只有一个

$$(\lambda - a)^k$$

**例 3.1.4**　已知 10 阶矩阵 $\boldsymbol{A}$ 的初级因子为

$$\lambda, \lambda, \lambda^2, \lambda - 1, (\lambda - 1)^2, (\lambda - 1)^3$$

求 $\boldsymbol{A}$ 的不变因子和行列式因子.

**解**　因为一次因式的最高次幂出现在 $d_{10}(\lambda)$ 中,所以

$$d_{10}(\lambda) = \lambda^2 (\lambda - 1)^3$$

剩下的初级因子中一次因式的最高次幂出现在 $d_9(\lambda)$ 中,所以

$$d_9(\lambda) = \lambda (\lambda - 1)^2$$

依次类推,可以得到各阶不变因子

$$d_8(\lambda) = \lambda(\lambda - 1)$$

$$d_7(\lambda) = d_6(\lambda) = \cdots = d_1(\lambda) = 1$$

根据 $d_1(\lambda) = D_1(\lambda), d_i(\lambda) = D_i(\lambda)/D_{i-1}(\lambda)\ (i = 2, \cdots, 10)$ 知各阶行列式因子为

$$D_1(\lambda) = \cdots = D_7(\lambda) = 1$$

$$D_8(\lambda) = \lambda(\lambda - 1)$$

$$D_9(\lambda) = \lambda^2(\lambda - 1)^3$$

$$D_7(\lambda) = \lambda^4(\lambda - 1)^6$$

现在考虑矩阵 $\boldsymbol{A}$ 的标准形问题.

设矩阵 $\boldsymbol{A} = (a_{ij})_n \in \mathbf{C}^{n\times n}$ 的全部初级因子为

$$(\lambda - \lambda_1)^{k_1}, (\lambda - \lambda_2)^{k_2}, \cdots, (\lambda - \lambda_t)^{k_t}$$

这里 $\lambda_1, \lambda_2, \cdots, \lambda_t$ 可能有相同的,指数 $k_1, k_2, \cdots, k_t$ 也可能有相同的,对每个初级因子 $(\lambda - \lambda_i)^{k_i}$ 构做一个 $k_i$ 阶矩阵(Jordan 块):

$$
\boldsymbol{J}_i = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix}, \quad i = 1, 2, \cdots, i
$$

由所有这些 Jordan 块构成的分块对角矩阵

$$
\boldsymbol{J} = \begin{pmatrix} \boldsymbol{J}_1 & & & \\ & \boldsymbol{J}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{J}_t \end{pmatrix}
$$

称为矩阵 $\boldsymbol{A}$ 的 Jordan 标准形.

**定理 3.1.1** 每个 $n$ 阶复矩阵 $\boldsymbol{A}$ 都与一个 Jordan 标准形 $\boldsymbol{J}$ 相似. 这个 Jordan 标准形在不计其中 Jordan 块的排列次序时,完全由矩阵 $\boldsymbol{A}$ 唯一确定. 即每个矩阵都有唯一的 Jordan 标准形.

定理 3.1.1 用线性变换的语言叙述就是:

设 $T$ 是复数域上 $n$ 维线性空间 $V$ 上的线性变换,则在 $V$ 中必存在一组基,使 $T$ 在这组基下的矩阵是 Jordan 形矩阵. 在不计 Jordan 块的排列次序时,这个 Jordan 形矩阵由 $T$ 唯一确定.

**推论 3.1.1** 复矩阵 $\boldsymbol{A}$ 与对角阵相似的充要条件是 $\boldsymbol{A}$ 的初级因子全是一次的.

由此可知,前面例 3.1.1、例 3.1.2 中矩阵的 Jordan 标准形分别为

$$
\boldsymbol{J} = \begin{pmatrix} 1 & & \\ & -2 & \\ & & -2 \end{pmatrix}
$$

与

$$
\boldsymbol{J} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}
$$

**例 3.1.5** 求矩阵

$$
\boldsymbol{A} = \begin{pmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{pmatrix}
$$

的 Jordan 标准形.

**解**　由 $\boldsymbol{A}(\lambda)=\lambda\boldsymbol{E}-\boldsymbol{A}$ 可以求得 $\boldsymbol{A}$ 的初级因子为 $\lambda-1$, $(\lambda-1)^2$, 故 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J}=\begin{pmatrix}1&0&0\\0&1&1\\0&0&1\end{pmatrix}$$

**例 3.1.6**　已知

$$\boldsymbol{A}=\begin{pmatrix}-1&-2&6\\-1&0&3\\-1&-1&4\end{pmatrix}$$

求可逆矩阵 $\boldsymbol{P}$,使 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$, 其中 $\boldsymbol{J}$ 为 $\boldsymbol{A}$ 的 Jordan 标准形.

**解**　由例 3.1.5 知 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J}=\begin{pmatrix}1&0&0\\0&1&1\\0&0&1\end{pmatrix}$$

记 $\boldsymbol{P}=(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3)$,由 $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J}$,有 $\boldsymbol{A}\boldsymbol{P}=\boldsymbol{P}\boldsymbol{J}$, 即

$$\boldsymbol{A}(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3)=(\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3)\begin{pmatrix}1&0&0\\0&1&1\\0&0&1\end{pmatrix}$$

比较上式两边有

$$\boldsymbol{A}\boldsymbol{p}_1=\boldsymbol{p}_1$$

$$\boldsymbol{A}\boldsymbol{p}_2=\boldsymbol{p}_2$$

$$\boldsymbol{A}\boldsymbol{p}_3=\boldsymbol{p}_2+\boldsymbol{p}_3$$

由方程组 $(\boldsymbol{E}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{0}$ 得到两个线性无关的特征向量

$$\boldsymbol{\xi}_1=(-1, 1, 0)^{\mathrm{T}},\qquad \boldsymbol{\xi}_2=(3, 0, 1)^{\mathrm{T}}$$

可以取

$$\boldsymbol{p}_1=\boldsymbol{\xi}_1=(-1, 1, 0)^{\mathrm{T}}$$

但不能简单地取 $\boldsymbol{p}_2=\boldsymbol{\xi}_2$,因为 $\boldsymbol{p}_2$ 的选取应该保证方程组$(\boldsymbol{E}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{p}_2$ 有解,因此,设 $\boldsymbol{p}_2=c_1\boldsymbol{\xi}_1+c_2\boldsymbol{\xi}_2$, 其中 $c_1$, $c_2$ 为待定系数. 从而方程组 $(\boldsymbol{E}-\boldsymbol{A})\boldsymbol{x}=\boldsymbol{p}_2$ 为

$$\begin{pmatrix} 2 & 2 & -6 \\ 1 & 1 & -3 \\ 1 & 1 & -3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = \begin{pmatrix} c_1 - 3c_2 \\ -c_1 \\ -c_2 \end{pmatrix}$$

易知，当 $c_1 = c_2$ 时方程组有解. 取 $c_1 = c_2 = 1$，得到

$$\boldsymbol{p}_2 = (2, 1, 1)^{\mathrm{T}}, \quad \boldsymbol{p}_3 = (2, 0, 1)^{\mathrm{T}}$$

于是，令

$$\boldsymbol{P} = (\boldsymbol{p}_1, \boldsymbol{p}_2, \boldsymbol{p}_3) = \begin{pmatrix} -1 & 2 & 2 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

则有
$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J}$$

**例 3.1.7** 求解线性微分方程组

$$\begin{cases} \dfrac{\mathrm{d}x_1}{\mathrm{d}t} = -x_1 - 2x_2 + 6x_3 \\ \dfrac{\mathrm{d}x_2}{\mathrm{d}t} = -x_1 + 3x_3 \\ \dfrac{\mathrm{d}x_3}{\mathrm{d}t} = -x_1 - x_2 + 4x_3 \end{cases}$$

**解** 记 $\boldsymbol{X} = (x_1, x_2, x_3)^{\mathrm{T}}$，$\dfrac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \left(\dfrac{\mathrm{d}x_1}{\mathrm{d}t}, \dfrac{\mathrm{d}x_2}{\mathrm{d}t}, \dfrac{\mathrm{d}x_3}{\mathrm{d}t}\right)^{\mathrm{T}}$，则方程组可以记为

$$\frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}t} = \boldsymbol{A}\boldsymbol{X}$$

其中

$$\boldsymbol{A} = \begin{pmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{pmatrix}$$

根据例 3.1.6 知，存在可逆阵

$$\boldsymbol{P} = \begin{pmatrix} -1 & 2 & 2 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}$$

使

$$P^{-1}AP = J = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

令 $X = PY$,其中 $Y = (y_1, y_2, y_3)^{\mathrm{T}}$, 则

$$APY = AX = \frac{\mathrm{d}X}{\mathrm{d}t} = P\frac{\mathrm{d}Y}{\mathrm{d}t}$$

所以

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = P^{-1}APY = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix}$$

即

$$\begin{cases} \dfrac{\mathrm{d}y_1}{\mathrm{d}t} = y_1 \\ \dfrac{\mathrm{d}y_2}{\mathrm{d}t} = y_2 + y_3 \\ \dfrac{\mathrm{d}y_3}{\mathrm{d}t} = y_3 \end{cases}$$

解之得

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \end{pmatrix} = \begin{pmatrix} c_1\mathrm{e}^t \\ (c_2 + c_3 t)\mathrm{e}^t \\ c_3\mathrm{e}^t \end{pmatrix}$$

所以

$$X = PY = \begin{pmatrix} -c_1 + 2c_2 + 2c_3(1+t) \\ c_1 + c_2 + c_3 t \\ c_2 + c_3(1+t) \end{pmatrix}\mathrm{e}^t$$

## 3.2 λ-矩阵及其 Smith 标准形

**定义 3.2.1** 形如

$$A(\lambda)=\begin{pmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ a_{21}(\lambda) & a_{22}(\lambda) & \cdots & a_{2n}(\lambda) \\ \vdots & \vdots & & \vdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots & a_{mn}(\lambda) \end{pmatrix}$$

的矩阵称为 $\lambda$ -矩阵或多项式矩阵. 其中元素 $a_{ij}(\lambda)(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 是 $\lambda$ 的多项式.

显然数字矩阵和特征矩阵 $\lambda E-A$ 都是 $\lambda$ -矩阵.

如同数字矩阵一样,可以定义 $\lambda$ -矩阵的相等、加法、数乘、乘法等,对于 $n$ 阶 $\lambda$ -矩阵可以定义行列式、子式、余子式、伴随矩阵等. 而 $\lambda$ -矩阵的秩定义为 $A(\lambda)$ 中不为零的子式的最大阶数,记为 $\mathrm{rank}\,A(\lambda)$ 或 $\mathrm{r}(A(\lambda))$. 当 $n$ 阶 $\lambda$ -矩阵的秩为 $n$ 时,称该 $\lambda$ -矩阵为满秩的或非奇异的. 矩阵 $A$ 的特征矩阵 $\lambda E-A$ 就是满秩 $\lambda$ -矩阵.

**定义 3.2.2** 若对于 $n$ 阶 $\lambda$ -矩阵有

$$A(\lambda)B(\lambda)=B(\lambda)A(\lambda)=E_n$$

则称 $A(\lambda)$ 可逆,称 $B(\lambda)$ 为 $A(\lambda)$ 的逆矩阵. 若 $A(\lambda)$ 有逆,则一定唯一. 与数字矩阵不同的是满秩矩阵 $A(\lambda)$ 不一定可逆.

**定理 3.2.1** $A(\lambda)$ 可逆的充要条件是 $A(\lambda)$ 的行列式为不等于零的常数,即

$$|A(\lambda)|=C\neq 0$$

**证** 设 $A(\lambda)$ 可逆,则有多项式矩阵 $B(\lambda)$,使

$$A(\lambda)B(\lambda)=B(\lambda)A(\lambda)=E_n$$

故有

$$|A(\lambda)||B(\lambda)|=|E_n|=1$$

故 $|A(\lambda)|$ 与 $|B(\lambda)|$ 只能是零次多项式. 所以当 $A(\lambda)$ 可逆时, $|A(\lambda)|$ 必定等于某个非零常数 $C$.

反之,若 $|A(\lambda)|=C\neq 0$, 易知 $A(\lambda)$ 可逆,且逆矩阵为 $\frac{1}{C}A^*(\lambda)$. 这里 $A^*(\lambda)$ 是 $A(\lambda)$ 的伴随矩阵. □

**例 3.2.1** 多项式矩阵

$$A(\lambda)=\begin{pmatrix} \lambda+1 & \lambda+3 \\ \lambda^2+3\lambda & \lambda^2+5\lambda+4 \end{pmatrix}$$

$$B(\lambda)=\begin{pmatrix} \lambda+1 & \lambda+3 \\ \lambda^2+3\lambda+2 & \lambda^2+5\lambda+6 \end{pmatrix}$$

中，$A(\lambda)$是可逆的，而$B(\lambda)$是不可逆的. 因为 $|A(\lambda)|=4$，$|B(\lambda)|=0$.

**定义 3.2.3**　$\lambda$-矩阵$A(\lambda)$的初等变换是指下面的三种变换：

(i) $A(\lambda)$的某两行(列)互换；

(ii) $A(\lambda)$的某行(列)乘非零常数 $k$；

(iii) $A(\lambda)$的某行(列)乘多项式 $\varphi(\lambda)$加到另一行(列)上去.

上述三种初等变换有相应的三种初等矩阵.

$$P(i,\ j),\quad P[i(k)],\quad P[j(\varphi),\ i]$$

且在施行初等行变换时左乘初等矩阵，施行初等列变换时右乘初等矩阵.

上面三种初等矩阵是满秩的. 满秩矩阵左(右)乘一个矩阵不改变原矩阵的秩.

**定义 3.2.4**　若 $\lambda$-矩阵 $A(\lambda)$经过有限次初等变换化为 $\lambda$-矩阵 $B(\lambda)$，则称 $A(\lambda)$ 与 $B(\lambda)$等价，记为$A(\lambda)\cong B(\lambda)$.

易证 $\lambda$-矩阵的这一等价定义满足下列性质：

(i) 反身性. $A(\lambda)\cong B(\lambda)$；

(ii) 对称性. 若$A(\lambda)\cong B(\lambda)$，则$B(\lambda)\cong A(\lambda)$；

(iii) 传递性. 若$A(\lambda)\cong B(\lambda)$，$B(\lambda)\cong C(\lambda)$，则$A(\lambda)\cong C(\lambda)$.

若两个 $\lambda$-矩阵等价，则它们的秩相等；反之，若两个 $\lambda$-矩阵的秩相等，但这两个 $\lambda$-矩阵未必等价，这与数字矩阵不同. 例如

$$A(\lambda)=\begin{pmatrix}\lambda & 1\\ 0 & \lambda\end{pmatrix},\qquad B(\lambda)=\begin{pmatrix}1 & -\lambda\\ 1 & \lambda\end{pmatrix}$$

它们的行列式分别为 $\lambda^2$ 和 $2\lambda$，显然秩都是 2. 但由初等变换的定义知，两个等价的 $\lambda$-矩阵的行列式只能相差一个不为零的常数因子，从而$A(\lambda)$与$B(\lambda)$不等价.

在 $\lambda$-矩阵的应用中，有多种标准形在不同场合中被使用. 这里我们只介绍其中最基本的一种，即 Smith(史密斯)标准形.

**定义 3.2.5**　形如下面的 $\lambda$-矩阵

$$J(\lambda)=\begin{bmatrix}d_1(\lambda) & 0 & \cdots & 0 & \cdots & 0\\ 0 & d_2(\lambda) & \cdots & 0 & \cdots & 0\\ \vdots & \vdots & & \vdots & & \vdots\\ 0 & 0 & \cdots & d_r(\lambda) & \cdots & 0\\ \vdots & \vdots & & \vdots & & \vdots\\ 0 & 0 & \cdots & 0 & \cdots & 0\end{bmatrix}$$

称为 Smith 标准形. 其中 $d_i(\lambda)(i=1,\ 2,\ \cdots,\ r)$ 为首一多项式(首项系数为 1)，并且 $d_i(\lambda)\mid d_{i+1}(\lambda)(i=1,\ 2,\ \cdots,\ r-1)$.

易知
$$\operatorname{rank} \boldsymbol{J}(\lambda) = r$$

下面研究如何将 λ-矩阵化为 Smith 标准形，进而得到求数字矩阵 $\boldsymbol{A}$ 的 Jordan 标准形的另一方法.

**引理 3.2.1** 设 λ-矩阵$\boldsymbol{A}(\lambda)$的左上角元素 $a_{11}(\lambda) \neq 0$，且$\boldsymbol{A}(\lambda)$中至少有一个元素不能被 $a_{11}(\lambda)$整除，则可以找到一个与$\boldsymbol{A}(\lambda)$等价的 λ-矩阵$\boldsymbol{B}(\lambda)$，其左上角元素 $b_{11}(\lambda) \neq 0$，且次数比 $a_{11}(\lambda)$的次数低.

**证** 分三种情况讨论：

(1) 若$\boldsymbol{A}(\lambda)$的第一列中有某个元素 $a_{i1}(\lambda)$不能被 $a_{11}(\lambda)$整除，则用 $a_{11}(\lambda)$去除 $a_{i1}(\lambda)$得

$$a_{i1}(\lambda) = q(\lambda)a_{11}(\lambda) + r(\lambda)$$

这里 $r(\lambda) \neq 0$，且次数比 $a_{11}(\lambda)$的次数低. 此时有

$$\boldsymbol{A}(\lambda) = \begin{bmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ \vdots & \vdots & & \vdots \\ a_{i1}(\lambda) & a_{i2}(\lambda) & \cdots & a_{in}(\lambda) \\ \vdots & \vdots & & \vdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots & a_{mn}(\lambda) \end{bmatrix}$$

$$\xrightarrow[\left[\begin{smallmatrix}\text{第一行乘}-q(\lambda)\\ \text{加到第 } i \text{ 行}\end{smallmatrix}\right]]{[(i)-q(\lambda)(1)]} \begin{bmatrix} a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ \vdots & \vdots & & \vdots \\ r(\lambda) & a_{i2}(\lambda)-q(\lambda)a_{12}(\lambda) & \cdots & a_{in}(\lambda)-q(\lambda)a_{1n}(\lambda) \\ \vdots & \vdots & & \vdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots & a_{mn}(\lambda) \end{bmatrix}$$

$$\xrightarrow[\text{[交换(1)行与}(i)\text{行]}]{[(1),\ (i)]} \begin{bmatrix} r(\lambda) & a_{i2}(\lambda)-q(\lambda)a_{12}(\lambda) & \cdots & a_{in}(\lambda)-q(\lambda)a_{1n}(\lambda) \\ \vdots & \vdots & & \vdots \\ a_{11}(\lambda) & a_{12}(\lambda) & \cdots & a_{1n}(\lambda) \\ \vdots & \vdots & & \vdots \\ a_{m1}(\lambda) & a_{m2}(\lambda) & \cdots & a_{mn}(\lambda) \end{bmatrix}$$

$$= \boldsymbol{B}(\lambda)$$

则 $\boldsymbol{B}(\lambda)$即为所求.

(2) 若$\boldsymbol{A}(\lambda)$的第一行中有某个元素 $a_{1j}(\lambda)$不能被 $a_{11}(\lambda)$整除，证法与(1)类似.

(3) 若$\boldsymbol{A}(\lambda)$中的第一行和第一列的所有元素都能被 $a_{11}(\lambda)$整除，但$\boldsymbol{A}(\lambda)$中至少有某个元素 $a_{ij}(\lambda)(i,\ j>1)$ 不能被 $a_{11}(\lambda)$整除，此时可以设

$$a_{i1}(\lambda) = \varphi(\lambda)a_{11}(\lambda)$$

将第一行乘以$-\varphi(\lambda)$加到第 $i$ 行上，得

$$\boldsymbol{A}(\lambda) \to \begin{bmatrix} a_{11}(\lambda) & \cdots & a_{1j}(\lambda) & \cdots & a_{1n}(\lambda) \\ \vdots & & \vdots & & \vdots \\ 0 & \cdots & a_{ij}(\lambda) - \varphi(\lambda)a_{1j}(\lambda) & \cdots & \\ \vdots & & \vdots & & \vdots \\ a_{m1}(\lambda) & \cdots & a_{mj}(\lambda) & \cdots & a_{mn}(\lambda) \end{bmatrix} = \boldsymbol{A}_1(\lambda)$$

再将 $\boldsymbol{A}_1(\lambda)$的第 $i$ 行加到第一行得

$$\boldsymbol{A}_1(\lambda) \to \begin{bmatrix} a_{11}(\lambda) & \cdots & a_{ij}(\lambda) + (1-\varphi(\lambda))a_{1j}(\lambda) & \cdots \\ \vdots & & \vdots & \\ 0 & \cdots & a_{ij}(\lambda) - \varphi(\lambda)a_{1j}(\lambda) & \cdots \\ \vdots & & \vdots & \\ a_{m1}(\lambda) & \cdots & a_{mj}(\lambda) & \cdots \end{bmatrix} = \boldsymbol{A}_2(\lambda)$$

则 $\boldsymbol{A}_2(\lambda)$的第一行元素中已至少有一个元素

$$a_{ij}(\lambda) + (1-\varphi(\lambda))a_{1j}(\lambda)$$

不能被 $a_{11}(\lambda)$整除，此时化为第(2)种情形，于是引理 3.2.1 得证. □

**定理 3.2.2**　若 $\boldsymbol{A}(\lambda)$是 $m \times n$ 的 $\lambda$ -矩阵，$\operatorname{rank}\boldsymbol{A}(\lambda) = r$，则 $\boldsymbol{A}(\lambda)$一定与 Smith 标准形

$$\boldsymbol{J}(\lambda) = \begin{bmatrix} d_1(\lambda) & & & & & \\ & \ddots & & & & \\ & & d_r(\lambda) & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \end{bmatrix}$$

等价.

**证**　若 $\operatorname{rank}\boldsymbol{A}(\lambda) = 0$，则 $\boldsymbol{A}(\lambda)$为零矩阵，无需讨论. 现设 $\operatorname{rank}\boldsymbol{A}(\lambda) = r > 0$，且 $\boldsymbol{A}(\lambda)$中 $a_{11}(\lambda) \neq 0$，若不然，可以通过行、列的变换做到这一点.

如果 $\boldsymbol{A}(\lambda)$的元素不是全部被 $a_{11}(\lambda)$整除，则由引理 3.2.1 可以找到一个等价矩阵，使其左上角元素的次数比 $a_{11}(\lambda)$低. 若这时左上角元素还不能整除矩阵的全部元素，则可以用同样的方法，逐步降低左上角元素的次数，直到得到一个等价矩阵 $\boldsymbol{B}(\lambda)$，其左上角元素 $b_{11}(\lambda)$可以整除其他所有元素，且 $b_{11}(\lambda)$为首一多项式. 由

于 $b_{11}(\lambda)$可以整除其他所有元素，可以将适当的多项式乘以第一行后加到其他各行，使第一列元素除 $b_{11}(\lambda)$外全部是零. 用同样的方法可以将第一行中除 $b_{11}(\lambda)$外的其他元素都化为零，这样得到与 $\boldsymbol{A}(\lambda)$等价的形如下式的矩阵：

$$\begin{bmatrix} d_1(\lambda) & 0 & 0 & \cdots & 0 \\ 0 & c_{22}(\lambda) & c_{23}(\lambda) & \cdots & c_{2n}(\lambda) \\ 0 & c_{32}(\lambda) & c_{33}(\lambda) & \cdots & c_{3n}(\lambda) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & c_{m2}(\lambda) & c_{m3}(\lambda) & \cdots & c_{mn}(\lambda) \end{bmatrix}$$

且 $d_1(\lambda)$(即 $b_{11}(\lambda)$)可以整除所有 $c_{ij}(\lambda)$.

现对上面 $(m-1)\times(n-1)$ 子阵$[c_{ij}(\lambda)]$重复上述过程，又可以得到与 $\boldsymbol{A}(\lambda)$等价的矩阵：

$$\begin{bmatrix} d_1(\lambda) & 0 & 0 & \cdots & 0 \\ 0 & d_2(\lambda) & 0 & \cdots & 0 \\ 0 & 0 & f_{33}(\lambda) & \cdots & f_{3n}(\lambda) \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & f_{m3}(\lambda) & \cdots & f_{mn}(\lambda) \end{bmatrix}.$$

这里
$$d_1(\lambda) \mid d_2(\lambda) \mid f_{ij}(\lambda) \ (i,\ j \geqslant 3)$$

这样一直做下去，$\boldsymbol{A}(\lambda)$最后化为所要求的标准形(即 Smith 标准形). □

**例 3.2.2** 试求 $\lambda$-矩阵

$$\boldsymbol{A}(\lambda) = \begin{bmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{bmatrix}$$

的 Smith 标准形.

**解** 对 $\boldsymbol{A}(\lambda)$进行初等变换，可得

$$\boldsymbol{A}(\lambda) = \begin{bmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{bmatrix}$$

$$\xrightarrow{c_3+c_1} \begin{bmatrix} 1 & \lambda^2 & \lambda \\ 0 & \lambda & -\lambda \\ 1 & \lambda^2 & -\lambda^2 \end{bmatrix}$$

$$\xrightarrow{r_3 - r_1} \begin{pmatrix} 1 & \lambda^2 & \lambda \\ 0 & \lambda & -\lambda \\ 0 & 0 & -\lambda^2 - \lambda \end{pmatrix}$$

$$\xrightarrow[-\lambda c_1 + c_3]{-\lambda^2 c_1 + c_2} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & -\lambda \\ 0 & 0 & -\lambda^2 - \lambda \end{pmatrix}$$

$$\xrightarrow{c_2 + c_3} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda(\lambda + 1) \end{pmatrix}$$

即为所求的 Smith 标准形. 且

$$d_1(\lambda) = 1, \quad d_2(\lambda) = \lambda, \quad d_3(\lambda) = \lambda(\lambda + 1)$$

下面讨论 $\lambda$-矩阵的一些性质. 与 $n$ 阶矩阵 $\boldsymbol{A}$ 的特征矩阵 $\lambda\boldsymbol{E} - \boldsymbol{A}$ 一样. 可以定义 $\lambda$-矩阵的行列式因子、不变因子与初级因子.

**定义 3.2.6** $\lambda$-矩阵 $\boldsymbol{A}(\lambda)$ 的所有 $k$ 阶子式的首一(最高次项系数为 1)最大公因式为 $\boldsymbol{A}(\lambda)$ 的 $k$ 阶行列式因子, 记为 $D_k(\lambda)(k = 1, 2, \cdots, n)$.

**定义 3.2.7** 设 $D_k(\lambda)$ 为 $n$ 阶 $\lambda$-矩阵 $\boldsymbol{A}(\lambda)$ 的 $k$ 阶行列式因子, 称

$$d_1(\lambda) = D_1(\lambda), \qquad d_k(\lambda) = \frac{D_k(\lambda)}{D_{k-1}(\lambda)}, \quad k = 2, \cdots, n$$

为 $\boldsymbol{A}(\lambda)$ 的不变因子.

**定义 3.2.8** 把 $\boldsymbol{A}(\lambda)$ 中的每个次数 $\geqslant 1$ 的不变因子 $d_k(\lambda)$ 分解为互不相同的一次因式方幂的乘积, 所有这些一次因式的方幂(相同的按出现的次数计算)称为 $\lambda$-矩阵 $\boldsymbol{A}(\lambda)$ 的初级因子.

**定理 3.2.3** 若 $\lambda$-矩阵 $\boldsymbol{A}(\lambda)$ 与 $\boldsymbol{B}(\lambda)$ 等价, 则它们具有相同的秩和相同的各级行列式因子.

**证** 只要证明经过一次初等变换后, 其秩与行列式因子不变即可.

设 $\lambda$-矩阵 $\boldsymbol{A}(\lambda)$ 经过一次初等变换变成 $\boldsymbol{B}(\lambda)$, $D_k(\lambda)$ 与 $T_k(\lambda)$ 分别为 $\boldsymbol{A}(\lambda)$ 与 $\boldsymbol{B}(\lambda)$ 的 $k$ 级行列式因子, 针对 3 种初等变换来证明 $D_k(\lambda) = T_k(\lambda)$.

(1) 变换 $\boldsymbol{A}(\lambda)$ 的某两行得到 $\boldsymbol{B}(\lambda)$, 这时 $\boldsymbol{B}(\lambda)$ 的每个 $k$ 阶子式或者等于 $\boldsymbol{A}(\lambda)$ 的某个 $k$ 阶子式, 或者是 $\boldsymbol{A}(\lambda)$ 的某个 $k$ 阶子式的 $-1$ 倍. $D_k(\lambda)$ 是 $\boldsymbol{B}(\lambda)$ 的 $k$ 阶子式的公因式, 从而 $D_k(\lambda) \mid T_k(\lambda)$.

(2) 用非零数 $a$ 乘 $\boldsymbol{A}(\lambda)$ 的某一行得到 $\boldsymbol{B}(\lambda)$, 这时 $\boldsymbol{B}(\lambda)$ 的每个 $k$ 阶子式或者等于 $\boldsymbol{A}(\lambda)$ 的某个 $k$ 阶子式, 或者等于 $\boldsymbol{A}(\lambda)$ 的某个 $k$ 阶子式的 $a$ 倍. 因此, $D_k(\lambda)$ 是

$\boldsymbol{B}(\lambda)$ 的 $k$ 阶子式的公因式，从而 $D_k(\lambda) \mid T_k(\lambda)$.

(3) 将 $\boldsymbol{A}(\lambda)$第 $j$ 行的 $\varphi(\lambda)$倍加到第 $i$ 行得到$\boldsymbol{B}(\lambda)$. 这时，$\boldsymbol{B}(\lambda)$中那些包含第 $i$ 行与第 $j$ 行的 $k$ 阶子式和那些不包含第 $i$ 行的 $k$ 阶子式都等于$\boldsymbol{A}(\lambda)$中对应的 $k$ 阶子式；$\boldsymbol{B}(\lambda)$中那些包含第 $i$ 行但不包含第 $j$ 行的 $k$ 阶子式等于$\boldsymbol{A}(\lambda)$中对应的一个 $k$ 阶子式与另一个 $k$ 阶子式的$\pm\varphi(\lambda)$倍之和，也就是$\boldsymbol{A}(\lambda)$的两个 $k$ 阶子式的组合. 因此，$D_k(\lambda)$是 $\boldsymbol{B}(\lambda)$的 $k$ 阶子式的公因式，从而 $D_k(\lambda) \mid T_k(\lambda)$.

由初等变换的可逆性，$\boldsymbol{B}(\lambda)$也可以经过一次初等行变换变成 $\boldsymbol{A}(\lambda)$. 由上面的讨论知，$T_k(\lambda) \mid D_k(\lambda)$. 所以有

$$D_k(\lambda) = T_k(\lambda)$$

对于初等列变换，可以类似地讨论.

当$\boldsymbol{A}(\lambda)$的全部 $k$ 阶子式为零时，$D_k(\lambda) = 0$，则 $T_k(\lambda) = 0$，$\boldsymbol{B}(\lambda)$的全部 $k$ 阶子式也全为零；反之亦然. 因此，$\boldsymbol{A}(\lambda)$与 $\boldsymbol{B}(\lambda)$既有相同的行列式因子，又有相同的秩. 结论得证. □

**定理 3.2.4** $\lambda$-矩阵$\boldsymbol{A}(\lambda)$的 Smith 标准形

$$\boldsymbol{J}(\lambda) = \begin{pmatrix} d_1(\lambda) & & & & & & \\ & d_2(\lambda) & & & & & \\ & & \ddots & & & & \\ & & & d_r(\lambda) & & & \\ & & & & 0 & & \\ & & & & & \ddots & \\ & & & & & & 0 \end{pmatrix}$$

是唯一的，且

$$d_1(\lambda) = D_1(\lambda), \qquad d_k(\lambda) = \frac{D_k(\lambda)}{D_{k-1}(\lambda)}, \quad k = 2, \cdots, r$$

**证** 因为$\boldsymbol{A}(\lambda)$与$\boldsymbol{J}(\lambda)$等价，由定理 3.2.3 知，它们有相同的行列式因子. 但$\boldsymbol{J}(\lambda)$为对角阵，且 $d_i(\lambda) \mid d_{i+1}(\lambda)$, $i = 1, 2, \cdots, r-1$; $d_i(\lambda) = 0$, $i > r$. 故

$$D_1(\lambda) = d_1(\lambda)$$

$$D_2(\lambda) = d_1(\lambda)d_2(\lambda)$$

$$\cdots\cdots$$

$$D_r(\lambda) = d_1(\lambda)d_2(\lambda)\cdots d_r(\lambda)$$

$$D_k(\lambda) = 0, \quad k > r$$

因此有

$$d_1(\lambda) = D_1(\lambda)$$

$$d_k(\lambda) = \frac{D_k(\lambda)}{D_{k-1}(\lambda)}, \quad k = 1, 2, \cdots, r-1$$

由于行列式因子在初等变换下不变,所以 $d_k(\lambda)$是唯一的. □

**定理 3.2.5** 两个 λ -矩阵$\boldsymbol{A}(\lambda)$与$\boldsymbol{B}(\lambda)$等价的充要条件是它们有相同的行列式因子(或相同的不变因子).

**证** 必要性由定理 3.2.3 已经得知;充分性证明如下:若$\boldsymbol{A}(\lambda)$, $\boldsymbol{B}(\lambda)$有相同的行列式因子,进而有相同的不变因子,则$\boldsymbol{A}(\lambda)$, $\boldsymbol{B}(\lambda)$与同一个 Smith 标准形等价,从而$\boldsymbol{A}(\lambda) \cong \boldsymbol{B}(\lambda)$. □

至此,实际上给出了求 λ -矩阵$\boldsymbol{A}(\lambda)$不变因子的一种方法,即将$\boldsymbol{A}(\lambda)$用初等变换化为 Smith 标准形. 则标准形对角线上元素 $d_k(\lambda)$ $(k = 1, 2, \cdots, r)$ 即为$\boldsymbol{A}(\lambda)$的不变因子.

数字矩阵相似的条件,也可以由其特征矩阵来描述. 应用上述 λ -矩阵的基础知识,可以推出下述结论.

**定理 3.2.6** 数域 $F$ 上两个 $n$ 阶矩阵$\boldsymbol{A}$, $\boldsymbol{B}$ 相似的充要条件是它们的特征矩阵$\lambda\boldsymbol{E} - \boldsymbol{A}$ 与$\lambda\boldsymbol{E} - \boldsymbol{B}$ 等价.

由定理 3.2.6 可知,$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 相似的充要条件是它们有相同的不变因子.

在复数域 $\mathbf{C}$ 上有:$\boldsymbol{A}$ 和 $\boldsymbol{B}$ 相似的充要条件是它们有相同的初级因子.

由此,我们得到求 $n$ 阶矩阵$\boldsymbol{A}$ 的初级因子的另一方法:即通过初等变换化 $\boldsymbol{A}$ 的特征矩阵$\lambda\boldsymbol{E} - \boldsymbol{A}$ 为 Smith 标准形,求出不变因式后,再计算出初级因子,从而可以得到与 $n$ 阶矩阵$\boldsymbol{A}$ 相似的 Jordan 标准形.

**例 3.2.3** 求矩阵

$$\boldsymbol{A} = \begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}$$

的 Jordan 标准形.

**解** 对 $\boldsymbol{A}$ 的特征矩阵$\boldsymbol{A}(\lambda) = \lambda\boldsymbol{E} - \boldsymbol{A}$ 进行初等变换,将其化为 Smith 标准形

$$\lambda\boldsymbol{E} - \boldsymbol{A} = \begin{pmatrix} \lambda+1 & -1 & 0 \\ 4 & \lambda-3 & 0 \\ -1 & 0 & \lambda-2 \end{pmatrix}$$

$$\xrightarrow{r_1 \leftrightarrow r_3} \begin{pmatrix} -1 & 0 & \lambda \\ 4 & \lambda-3 & 0 \\ \lambda+1 & -1 & 0 \end{pmatrix}$$

$$\xrightarrow[r_3+(\lambda+1)r_1]{r_2+4r_1} \begin{pmatrix} -1 & 0 & \lambda-2 \\ 0 & \lambda-3 & 4\lambda-8 \\ 0 & -1 & (\lambda+1)(\lambda-2) \end{pmatrix}$$

$$\xrightarrow{r_2 \leftrightarrow r_3} \begin{pmatrix} -1 & 0 & \lambda-2 \\ 0 & -1 & (\lambda+1)(\lambda-2) \\ 0 & \lambda-3 & 4\lambda-8 \end{pmatrix}$$

$$\xrightarrow{r_3+(\lambda-3)r_2} \begin{pmatrix} -1 & 0 & \lambda-2 \\ 0 & -1 & (\lambda+1)(\lambda-2) \\ 0 & 0 & (\lambda-2)(\lambda-1)^2 \end{pmatrix}$$

$$\longrightarrow \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda-2)(\lambda-1)^2 \end{pmatrix}$$

所以 $\boldsymbol{A}$ 的不变因子为

$$d_1(\lambda) = d_2(\lambda) = 1, \quad d_3(\lambda) = (\lambda-2)(\lambda-1)^2$$

从而其初级因子为 $(\lambda-2)$,$(\lambda-1)^2$，因此 $\boldsymbol{A}$ 的 Jordan 标准形为

$$\boldsymbol{J} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$$

## 3.3 Cayley-Hamilton 定理与矩阵的最小多项式

在线性代数中，我们给出了 $n$ 阶矩阵 $\boldsymbol{A}$ 的特征多项式，本节将进一步给出特征多项式的性质，其中最重要的就是 Cayley-Hamilton(凯莱-哈密顿)定理. 我们还将讨论矩阵的最小多项式及其与特征多项式的关系，进而通过矩阵 $\boldsymbol{A}$ 的特征多项式和最小多项式来分析 $\boldsymbol{A}$ 的 Jordan 标准形.

设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$，其特征多项式为

$$f(\lambda) = |\lambda \boldsymbol{E} - \boldsymbol{A}| = \lambda^n + a_1\lambda^{n-1} + a_2\lambda^{n-2} + \cdots + a_{n-1}\lambda + a_n$$

矩阵 $\boldsymbol{A}$ 与其特征多项式之间有如下重要关系：

**定理 3.3.1** (Cayley-Hamilton 定理)设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$, $f(\lambda) = |\lambda \boldsymbol{E} - \boldsymbol{A}|$，则 $f(\boldsymbol{A}) = \boldsymbol{0}$.

**证** 设 $\boldsymbol{J}$ 为 $\boldsymbol{A}$ 的 Jordan 标准形,则存在可逆矩阵 $\boldsymbol{P}$,使

$$\boldsymbol{P}^{-1}\boldsymbol{AP}=\boldsymbol{J}=\begin{pmatrix}\lambda_1 & k_1 & & & \\ & \lambda_2 & k_2 & & \\ & & \lambda_3 & \ddots & \\ & & & \ddots & k_{n-1} \\ & & & & \lambda_n\end{pmatrix}$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 为 $\boldsymbol{A}$ 的特征值,它们之中可以有相同者;$k_i$ 为 1 或 0,$i=1, 2, \cdots, n-1$.

$$\boldsymbol{A}=\boldsymbol{PJP}^{-1}$$

于是

$$f(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-\lambda_1)(\lambda-\lambda_2)\cdots(\lambda-\lambda_n)$$

从而

$$\begin{aligned}f(\boldsymbol{A})&=(\boldsymbol{A}-\lambda_1\boldsymbol{E})(\boldsymbol{A}-\lambda_2\boldsymbol{E})\cdots(\boldsymbol{A}-\lambda_n\boldsymbol{E})\\&=(\boldsymbol{PJP}^{-1}-\lambda_1\boldsymbol{E})(\boldsymbol{PJP}^{-1}-\lambda_2\boldsymbol{E})\cdots(\boldsymbol{PJP}^{-1}-\lambda_n\boldsymbol{E})\\&=\boldsymbol{P}[(\boldsymbol{J}-\lambda_1\boldsymbol{E})(\boldsymbol{J}-\lambda_2\boldsymbol{E})\cdots(\boldsymbol{J}-\lambda_n\boldsymbol{E})]\boldsymbol{P}^{-1}\\&=\boldsymbol{P}\left(\begin{pmatrix}0 & k_1 & & & \\ & \lambda_2-\lambda_1 & k_2 & & \\ & & \lambda_3-\lambda_1 & \ddots & \\ & & & \ddots & k_{n-1} \\ & & & & \lambda_n-\lambda_1\end{pmatrix}\right.\\&\quad\begin{pmatrix}\lambda_1-\lambda_2 & k_1 & & & \\ & 0 & k_2 & & \\ & & \lambda_3-\lambda_2 & \ddots & \\ & & & \ddots & k_{n-1} \\ & & & & \lambda_n-\lambda_2\end{pmatrix}\\&\quad\left.\cdots\begin{pmatrix}\lambda_1-\lambda_n & k_1 & & & \\ & \lambda_2-\lambda_n & k_2 & & \\ & & \lambda_3-\lambda_n & \ddots & \\ & & & \ddots & k_{n-1} \\ & & & & 0\end{pmatrix}\right)\boldsymbol{P}^{-1}\end{aligned}$$

$$
= P\left(\begin{pmatrix} 0 & 0 & * & \cdots & * \\ 0 & 0 & * & \cdots & * \\ \vdots & \vdots & * & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & & & * \end{pmatrix}\begin{pmatrix} \lambda_1-\lambda_3 & k_1 & & & \\ & \lambda_2-\lambda_3 & k_2 & & \\ & & 0 & \ddots & \\ & & & \ddots & k_{n-1} \\ & & & & \lambda_n-\lambda_3 \end{pmatrix}\right.
$$

$$
\left.\cdots\begin{pmatrix} \lambda_1-\lambda_n & k_1 & & & \\ & \lambda_2-\lambda_n & k_2 & & \\ & & \lambda_3-\lambda_n & \ddots & \\ & & & \ddots & k_{n-1} \\ & & & & 0 \end{pmatrix}\right)P^{-1} = \mathbf{0}
$$

由 Cayley-Hamilton 定理可以简化矩阵计算.

**例 3.3.1** 设

$$
A = \begin{pmatrix} -1 & 1 & 0 \\ -4 & 3 & 0 \\ 1 & 0 & 2 \end{pmatrix}
$$

试计算 $\varphi(A) = A^5 - 4A^4 + 6A^3 - 6A^2 + 6A - 3E$.

**解** $A$ 的特征多项式为

$$
f(\lambda) = |\lambda E - A| = \lambda^3 - 4\lambda^2 + 5\lambda - 2
$$

令

$$
\varphi(\lambda) = \lambda^5 - 4\lambda^4 + 6\lambda^3 - 6\lambda^2 + 6\lambda - 3
$$

易求得

$$
\varphi(\lambda) = (\lambda^2 + 1)f(\lambda) + \lambda - 1
$$

由于 $f(A) = 0$，故

$$
\varphi(A) = A - E = \begin{pmatrix} -2 & 1 & 0 \\ -4 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}
$$

**定义 3.3.1** 设 $A \in \mathbf{C}^{n\times n}$，$\varphi(\lambda)$是 $\lambda$ 的多项式，若 $\varphi(A) = \mathbf{0}$，则称 $\varphi(\lambda)$是 $A$ 的零化多项式.

由 Cayley-Hamilton 定理知，$A$ 的特征多项式是 $A$ 的零化多项式，故 $n$ 阶矩阵 $A$ 的零化多项式一定存在.

显然,若 $\varphi(\lambda)$是 $\boldsymbol{A}$ 的零化多项式,则对任意多项式 $g(\lambda)$, $\varphi(\lambda)g(\lambda)$也是 $\boldsymbol{A}$ 的零化多项式,可见 $\boldsymbol{A}$ 的零化多项式无最高次数者,因此我们所关心的是 $\boldsymbol{A}$ 的次数最低的零化多项式.

**定义 3.3.2** 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,在 $\boldsymbol{A}$ 的零化多项式中,次数最低的首一多项式称为 $\boldsymbol{A}$ 的最小多项式,记为 $m_A(\lambda)$.

**定理 3.3.2** $\boldsymbol{A}$ 的最小多项式 $m_A(\lambda)$整除 $\boldsymbol{A}$ 的任一零化多项式,特别地,$m_A(\lambda)\mid f(\lambda)$,其中 $f(\lambda)$为 $\boldsymbol{A}$ 的特征多项式.

**证** 设 $\varphi(\lambda)$是 $\boldsymbol{A}$ 的任一零化多项式,又 $m_A(\lambda)$是 $\boldsymbol{A}$ 的最小多项式,以 $m_A(\lambda)$除 $\varphi(\lambda)$得

$$\varphi(\lambda)=q(\lambda)m_A(\lambda)+\gamma(\lambda)$$

这里 $\gamma(\lambda)$若不为零,其次数小于 $m_A(\lambda)$的次数. 于是有

$$\varphi(\boldsymbol{A})=q(\boldsymbol{A})m_A(\boldsymbol{A})+\gamma(\boldsymbol{A})$$

因
$$\varphi(\boldsymbol{A})=m_A(\boldsymbol{A})=0$$

所以
$$\gamma(\boldsymbol{A})=0$$

即 $\gamma(\lambda)$也是 $\boldsymbol{A}$ 的零化多项式. 如果 $\gamma(\lambda)\neq 0$, 则 $\gamma(\lambda)$的次数小于 $m_A(\lambda)$的次数,这与 $m_A(\lambda)$是最小多项式矛盾. 所以 $\gamma(\lambda)=0$, 即 $m_A(\lambda)\mid\varphi(\lambda)$. 又 $f(\boldsymbol{A})=\boldsymbol{0}$. 所以 $m_A(\lambda)\mid f(\lambda)$. 即 $\boldsymbol{A}$ 的最小多项式一定是其特征多项式的因式. □

**定理 3.3.3** 矩阵 $\boldsymbol{A}$ 的最小多项式是唯一的.

**证** 若 $m_A(\lambda)$与 $n_A(\lambda)$均为 $\boldsymbol{A}$ 的最小多项式,则每一个都可以被另一个整除,因此两者只有常数因子的差别. 又两者都是首一多项式,这个常数因子必定等于1,所以 $m_A(\lambda)=n_A(\lambda)$.

**定理 3.3.4** 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,矩阵 $\boldsymbol{A}$ 的最小多项式的根必定是 $\boldsymbol{A}$ 的特征根;反之,$\boldsymbol{A}$ 的特征根也必定是 $\boldsymbol{A}$ 的最小多项式的根.

**证** 由定理 3.3.2 知,$\boldsymbol{A}$ 的最小多项式 $m_A(\lambda)$是 $\boldsymbol{A}$ 的特征多项式 $f(\lambda)$的因式,所以,$m_A(\lambda)$的根都是 $f(\lambda)$的根.

反之,若 $\lambda_0$ 是 $\boldsymbol{A}$ 的一个特征根,且

$$\boldsymbol{A\alpha}=\lambda_0\boldsymbol{\alpha},\quad \boldsymbol{\alpha}\neq\boldsymbol{0}$$

又设 $\boldsymbol{A}$ 的最小多项式为

$$m_A(\lambda)=\lambda^k+b_1\lambda^{k-1}+\cdots+b_{k-1}\lambda+b_k$$

则

$$m_A(\boldsymbol{A})\boldsymbol{\alpha}=\boldsymbol{A}^k\boldsymbol{\alpha}+b_1\boldsymbol{A}^{k-1}\boldsymbol{\alpha}+\cdots+b_{k-1}\boldsymbol{A\alpha}+b_k\boldsymbol{\alpha}$$

$$\begin{aligned}
&= \lambda_0^k \boldsymbol{\alpha} + b_1\lambda_0^{k-1}\boldsymbol{\alpha} + \cdots + b_{k-1}\lambda_0\boldsymbol{\alpha} + b_k\boldsymbol{\alpha} \\
&= (\lambda_0^k + b_1\lambda_0^{k-1} + \cdots + b_{k-1}\lambda_0 + b_k)\boldsymbol{\alpha} \\
&= m_A(\lambda_0)\boldsymbol{\alpha}
\end{aligned}$$

由于 $m_A(\boldsymbol{A}) = 0$，又 $\boldsymbol{\alpha} \neq \boldsymbol{0}$，所以 $m_A(\lambda_0) = 0$，亦即 $\lambda_0$ 是 $m_A(\lambda)$ 的根. □

定理 3.3.4 反映了矩阵 $\boldsymbol{A}$ 的特征多项式与最小多项式之间的重要关系. 由此可以得到求最小多项式的一个方法.

设矩阵 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$ 的所有不同的特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_t$，又 $\boldsymbol{A}$ 的特征多项式为

$$f(\lambda) = |\lambda\boldsymbol{E} - \boldsymbol{A}| = (\lambda - \lambda_1)^{s_1}(\lambda - \lambda_2)^{s_2}\cdots(\lambda - \lambda_t)^{s_t}$$

则 $\boldsymbol{A}$ 的最小多项式必具有如下形式

$$m_A(\lambda) = (\lambda - \lambda_1)^{k_1}(\lambda - \lambda_2)^{k_2}\cdots(\lambda - \lambda_t)^{k_t}$$

这里每个 $k_i \leqslant s_i$，$i = 1, 2, \cdots, t$.

**例 3.3.2** 求矩阵

$$\boldsymbol{A} = \begin{pmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 2 \end{pmatrix}$$

的最小多项式 $m_A(\lambda)$.

**解** $\boldsymbol{A}$ 的特征多项式为

$$f(\lambda) = |\lambda\boldsymbol{E} - \boldsymbol{A}| = (\lambda - 2)^2(\lambda - 3)$$

故 $\boldsymbol{A}$ 的最小多项式只能是

$$(\lambda - 2)(\lambda - 3) \quad \text{或} \quad (\lambda - 2)^2(\lambda - 3)$$

但由 $(\boldsymbol{A} - 2\boldsymbol{E})(\boldsymbol{A} - 3\boldsymbol{E}) = \boldsymbol{0}$ 知 $\boldsymbol{A}$ 的最小多项式为

$$m_A(\lambda) = (\lambda - 2)(\lambda - 3)$$

如果矩阵 $\boldsymbol{A}$ 与矩阵 $\boldsymbol{B}$ 相似，则由相似的性质和定理 3.3.4 知，$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 有相同的最小多项式. 但是要注意，这个条件并不是充分的，即最小多项式相同的矩阵不一定是相似的. 下面的例子说明了这个结论.

**例 3.3.3** 设

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}, \quad \boldsymbol{B} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}$$

$\boldsymbol{A}$ 与 $\boldsymbol{B}$ 的最小多项式都是 $(\lambda-1)^2(\lambda-2)$，但它们的特征多项式不同，因此 $\boldsymbol{A}$ 和 $\boldsymbol{B}$ 不是相似的.

**例 3.3.4**　试求 $k$ 级 Jordan 块

$$\boldsymbol{J}=\begin{pmatrix} a & 1 & & & \\ & a & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & a \end{pmatrix}$$

的最小多项式.

**解**　$\boldsymbol{J}$ 的特征多项式为 $(x-a)^k$，而

$$\boldsymbol{J}-a\boldsymbol{E}=\begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ 0 & 0 & 0 & \cdots & 0 \end{pmatrix}$$

$$(\boldsymbol{J}-a\boldsymbol{E})^{k-1}=\begin{pmatrix} 0 & \cdots & 0 & 1 \\ 0 & \cdots & 0 & 0 \\ \vdots & & \vdots & \vdots \\ 0 & \cdots & 0 & 0 \end{pmatrix}\neq \boldsymbol{0}$$

所以 $\boldsymbol{J}$ 的最小多项式为 $(x-a)^k$.

**例 3.3.5**　试求矩阵

$$\boldsymbol{A}=\begin{pmatrix} -2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 3 \end{pmatrix}$$

的最小多项式.

**解**　$\boldsymbol{A}$ 的特征多项式是 $(\lambda+2)^5(\lambda-3)^3$，故 $\boldsymbol{A}$ 的最小多项式的形式是 $(\lambda+2)^k(\lambda-3)^l$.

由于

$$A+2E=\begin{pmatrix}0&1&&&&&&\\0&0&&&&&&\\&&0&1&0&&&\\&&0&0&1&&&\\&&0&0&0&&&\\&&&&&5&&\\&&&&&&5&1\\&&&&&&0&5\end{pmatrix}$$

$$A-3E=\begin{pmatrix}-5&1&&&&&&\\0&-5&&&&&&\\&&-5&1&0&&&\\&&0&-5&1&&&\\&&0&0&-5&&&\\&&&&&0&&\\&&&&&&0&1\\&&&&&&0&0\end{pmatrix}$$

所以由分块矩阵乘法可知,要使 $(A+2E)^k(A-3E)^l=0$,只需注意使幂零矩阵的那部分变为零矩阵所需乘积的次数. $A+2E$ 需自乘 3 次,即在 $(A+2E)^3$ 中原幂零矩阵的那部分变为零矩阵,而 $A-3E$ 只需自乘 2 次便可以达到目的. 因此,$A$ 的最小多项式是

$$m_A(\lambda)=(\lambda+2)^3(\lambda-3)^2$$

一般地,如果 $A$ 是一个准对角阵

$$A=\begin{pmatrix}A_1&0\\0&A_2\end{pmatrix}$$

且 $A_1$ 的最小多项式为 $m_{A_1}(\lambda)$,$A_2$ 的最小多项式为 $m_{A_2}(\lambda)$,那么 $A$ 的最小多项式为 $m_{A_1}(\lambda)$ 与 $m_{A_2}(\lambda)$ 的最小公倍式$[m_{A_1}(\lambda),\ m_{A_2}(\lambda)]$.

事实上,记

$$g(\lambda)=[m_{A_1}(\lambda),\ m_{A_2}(\lambda)]$$

首先

$$g(A)=\begin{pmatrix}g(A_1)&0\\0&g(A_2)\end{pmatrix}=0$$

因此 $g(\lambda)$能被 $\boldsymbol{A}$ 的最小多项式整除. 其次,如果 $h(\lambda)$是 $\boldsymbol{A}$ 的最小多项式,那么

$$h(\boldsymbol{A}) = \begin{pmatrix} h(\boldsymbol{A}_1) & 0 \\ 0 & h(\boldsymbol{A}_2) \end{pmatrix} = \boldsymbol{0}$$

所以

$$h(\boldsymbol{A}_1) = 0,\ h(\boldsymbol{A}_2) = 0$$

因而

$$m_{\boldsymbol{A}_1}(\lambda) \mid h(\lambda), \qquad m_{\boldsymbol{A}_2}(\lambda) \mid h(\lambda)$$

由此知 $g(\lambda) \mid h(\lambda)$. 从而 $g(\lambda)$是 $\boldsymbol{A}$ 的最小多项式,即

$$m_{\boldsymbol{A}}(\lambda) = g(\lambda) = [m_{\boldsymbol{A}_1}(\lambda),\ m_{\boldsymbol{A}_2}(\lambda)]$$

这个结论可以推广到 $\boldsymbol{A}$ 为若干个矩阵组成的准对角阵的情形. 即如果

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_1 & & & \\ & \boldsymbol{A}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{A}_s \end{pmatrix}$$

$\boldsymbol{A}_i$ 的最小多项式为 $m_{\boldsymbol{A}_i}(\lambda)(i = 1,\ 2,\ \cdots,\ s)$, 则 $\boldsymbol{A}$ 的最小多项式为

$$m_{\boldsymbol{A}}(\lambda) = [m_{\boldsymbol{A}_1}(\lambda),\ m_{\boldsymbol{A}_2}(\lambda),\ \cdots,\ m_{\boldsymbol{A}_s}(\lambda)]$$

特别地,$\boldsymbol{A}$ 的 Jordan 标准形的最小多项式为

$$m_{\boldsymbol{J}}(\lambda) = m_{\boldsymbol{A}}(\lambda) = (\lambda - \lambda_1)^{q_1} \cdots (\lambda - \lambda_i)^{q_i} \cdots (\lambda - \lambda_s)^{q_s}$$

其中,$\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_s$ 为 $\boldsymbol{A}$ 的 $s$ 个不同的特征值;$q_i$ 是 $\boldsymbol{A}$ 的 Jordan 标准形的 Jordan 块 $\boldsymbol{J}_i$ 子块 $\boldsymbol{J}_{i1},\ \boldsymbol{J}_{i2},\ \cdots,\ \boldsymbol{J}_{ir_1}$ 中阶数最高者, $i = 1,\ 2,\ \cdots,\ s$.

例如, $\boldsymbol{A} = \begin{pmatrix} -1 & -2 & 6 \\ -1 & 0 & 3 \\ -1 & -1 & 4 \end{pmatrix}$ 的 Jordan 标准形为 $\boldsymbol{J} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}$, 则由上述内容可知,$\boldsymbol{A}$ 的最小多项式为

$$m_{\boldsymbol{A}}(\lambda) = (\lambda - 1)^2$$

综上分析有下述定理.

**定理 3.3.5**　设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$, 则 $\boldsymbol{A}$ 的最小多项式为 $\boldsymbol{A}$ 的第 $n$ 个不变因子 $d_n(\lambda)$.

**定理 3.3.6**　设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$, 则 $\boldsymbol{A}$ 相似于对角阵的充分必要条件是 $\boldsymbol{A}$ 的最小多项式没有重根.

## 习题3

1. 试求下列矩阵的不变因子与初级因子：

(1) $A=\begin{pmatrix}-1 & & & \\ & -2 & & \\ & & 1 & \\ & & & 2\end{pmatrix}$；

(2) $A=\begin{pmatrix}a & -b_1 & & & & \\ & a & -b_2 & & & \\ & & \ddots & \ddots & & \\ & & & a & -b_i & \\ & & & & \ddots & \ddots \\ & & & & \ddots & -b_{n-1} \\ & & & & & a\end{pmatrix}$ $(b_i\neq 0, i=1,2,\cdots,n-1)$；

(3) $A=\begin{pmatrix}-1 & 1 & 0\\ -4 & 3 & 0\\ 1 & 0 & 2\end{pmatrix}$.

2. 试求下列矩阵的 Smith 标准形.

(1) $A(\lambda)=\begin{pmatrix}\lambda-2 & -1 & 0\\ 0 & \lambda-2 & -1\\ 0 & 0 & \lambda-2\end{pmatrix}$；

(2) $A(\lambda)=\begin{pmatrix}\lambda^2+\lambda & 0 & 0\\ 0 & \lambda & 0\\ 0 & 0 & (\lambda+1)^2\end{pmatrix}$.

3. 利用 Cayley-Hamilton 定理证明：任意可逆矩阵 $A$ 的逆矩阵 $A^{-1}$ 都可以表示为 $A$ 的多项式.

4. 设 $A=\begin{pmatrix}1 & -1\\ 2 & 5\end{pmatrix}$. 证明：$B=2A^4-12A^3+19A^2-29A+37E$ 是可逆矩阵，并把 $B^{-1}$ 表示成 $A$ 的多项式.

5. 设 $A=\begin{pmatrix}1 & 0 & 2\\ 0 & -1 & 1\\ 0 & 1 & 0\end{pmatrix}$，试由 Cayley-Hamilton 定理计算 $g(A)=2A^8-3A^5+A^4+A^2-4E$.

6. 试在复数域内，求下列矩阵的 Jordan 标准形：

(1) $\boldsymbol{A}=\begin{pmatrix}1 & -1 & 2\\ 3 & -3 & 6\\ 2 & -2 & 4\end{pmatrix}$；

(2) $\boldsymbol{A}=\begin{pmatrix}3 & 7 & -3\\ -2 & -5 & 2\\ -4 & -10 & 3\end{pmatrix}$；

(3) $\boldsymbol{A}=\begin{pmatrix}3 & 0 & 8\\ 3 & -1 & 6\\ -2 & 0 & -5\end{pmatrix}$；

(4) $\boldsymbol{A}=\begin{pmatrix}4 & 5 & -2\\ -2 & -2 & 1\\ -1 & -1 & 1\end{pmatrix}$；

(5) $\boldsymbol{A}=\begin{pmatrix}3 & 1 & & \\ -4 & -1 & & \\ & & 2 & 1\\ & & -1 & 0\end{pmatrix}$.

7. 试用矩阵的 Jordan 标准形证明 $\boldsymbol{A}$ 与 $\boldsymbol{A}^{\mathrm{T}}$ 相似.

8. 设 $\boldsymbol{A}=\begin{pmatrix}-1 & 0 & 1\\ 1 & 2 & 0\\ -4 & 0 & 3\end{pmatrix}$. 试求 $\boldsymbol{A}^k$.

9. 试应用矩阵的 Jordan 标准形求解线性微分方程组

$$\begin{cases}\dfrac{\mathrm{d}x_1}{\mathrm{d}t}=2x_1+2x_2-x_3\\[2mm] \dfrac{\mathrm{d}x_2}{\mathrm{d}t}=-x_1-x_2+x_3\\[2mm] \dfrac{\mathrm{d}x_3}{\mathrm{d}t}=-x_1-2x_2+2x_3\end{cases}$$

这里 $x_1$，$x_2$，$x_3$ 都是 $t$ 的未知函数.

10. 求 $\boldsymbol{A}=\begin{pmatrix}2 & 2 & -1\\ -1 & -1 & 1\\ -1 & -2 & 2\end{pmatrix}$ 的最小多项式.

11. 证明：若 $n$ 阶方阵满足 $\boldsymbol{A}^2-5\boldsymbol{A}=-6\boldsymbol{E}$，则 $\boldsymbol{A}$ 相似于对角阵.

12. 设矩阵 $\boldsymbol{A}$ 的特征多项式为 $f(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-2)^2(\lambda-3)^2$.

(1) 给出 $\boldsymbol{A}$ 的所有可能的最小多项式;

(2) 给出 $\boldsymbol{A}$ 的所有可能的 Jordan 矩阵.

# 第4章 矩阵分解

将矩阵表示成特定类型矩阵的乘积，这种表示称为矩阵的分解.矩阵的分解理论与方法是矩阵分析理论的重要内容，也在解线性方程组，求特征值，求广义逆矩阵的实际计算中有着重要的应用价值.

## 4.1 矩阵的LU分解

**定义 4.1.1** 设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$，若 $\boldsymbol{A}$ 可以表示成一个下三角矩阵 $\boldsymbol{L}$ 与一个上三角矩阵 $\boldsymbol{U}$ 的乘积

$$\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$$

则称其为矩阵 $\boldsymbol{A}$ 的 LU 分解(三角分解).

**定理 4.1.1** 设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$，如果 $\boldsymbol{A}$ 的顺序主子式

$$a_{11} \neq 0,\quad \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} \neq 0,\ \cdots,\ \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} \\ \vdots & \vdots & & \vdots \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} \end{vmatrix} \neq 0$$

则存在唯一的主对角线上元素全为1的下三角阵 $\boldsymbol{L}$ 与唯一的上三角矩阵 $\boldsymbol{U}$，使得 $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$.

**证** 我们对矩阵 $\boldsymbol{A}$ 的阶数 $n$ 使用数学归纳法.明显地，当 $n=1$ 时，$a_{11} = 1 \cdot a_{11}$ 就是唯一的分解式.

现在假定对 $n-1$ 阶矩阵，定理的结论成立.我们证明对 $n$ 阶矩阵结论成立.

对 $\boldsymbol{A}$ 进行分块

$$\boldsymbol{A} = \begin{pmatrix} \boldsymbol{A}_{n-1} & \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2^{\mathrm{T}} & a_{nn} \end{pmatrix}$$

其中 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2 \in \mathbf{C}^{n-1}$. 由于 $n-1$ 阶矩阵 $\boldsymbol{A}_{n-1}$ 的 $k$ 阶顺序主子式就是 $\boldsymbol{A}$ 的 $k$ 阶顺序主子式 $(k=1, 2, \cdots, n-2)$，故它们都不为零. 从而由归纳假设，$\boldsymbol{A}_{n-1}$ 有唯一的 LU 分解

$$\boldsymbol{A}_{n-1} = \boldsymbol{L}_{n-1}\boldsymbol{U}_{n-1}$$

其中 $\boldsymbol{L}_{n-1}$ 的主对角线上的元都是 1. 由于

$$|\boldsymbol{A}_{n-1}| = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1,n-1} \\ a_{21} & a_{22} & \cdots & a_{2,n-1} \\ \vdots & \vdots & & \vdots \\ a_{n-1,1} & a_{n-1,2} & \cdots & a_{n-1,n-1} \end{vmatrix} = |\boldsymbol{L}_{n-1}|\,|\boldsymbol{U}_{n-1}| \neq 0$$

所以，$\boldsymbol{L}_{n-1}$ 及 $\boldsymbol{U}_{n-1}$ 是 $n-1$ 阶可逆矩阵.

我们先假设已有 $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$，其中

$$\boldsymbol{L} = \begin{pmatrix} \boldsymbol{L}_{n-1} & \boldsymbol{0} \\ \boldsymbol{\gamma}^{\mathrm{T}} & 1 \end{pmatrix}, \qquad \boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_{n-1} & \boldsymbol{\beta} \\ \boldsymbol{0}^{\mathrm{T}} & u_{nn} \end{pmatrix}$$

这里 $\boldsymbol{\beta}, \boldsymbol{\gamma} \in \mathbf{C}^{n-1}$，$u_{nn}$ 待定. 作乘积

$$\boldsymbol{L}\boldsymbol{U} = \begin{pmatrix} \boldsymbol{L}_{n-1}\boldsymbol{U}_{n-1} & \boldsymbol{L}_{n-1}\boldsymbol{\beta} \\ \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{U}_{n-1} & \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\beta} + u_{nn} \end{pmatrix} = \begin{pmatrix} \boldsymbol{A}_{n-1} & \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2^{\mathrm{T}} & a_{nn} \end{pmatrix} = \boldsymbol{A}$$

则 $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $u_{nn}$ 必须满足

$$\boldsymbol{L}_{n-1}\boldsymbol{\beta} = \boldsymbol{\alpha}_1, \quad \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{U}_{n-1} = \boldsymbol{\alpha}_2^{\mathrm{T}}, \quad \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\beta} + u_{nn} = a_{nn}$$

注意到 $\boldsymbol{L}_{n-1}$, $\boldsymbol{U}_{n-1}$ 可逆，所以由上式可以唯一确定

$$\boldsymbol{\beta} = \boldsymbol{L}_{n-1}^{-1}\boldsymbol{\alpha}_1, \quad \boldsymbol{\gamma}^{\mathrm{T}} = \boldsymbol{\alpha}_2^{\mathrm{T}}\boldsymbol{U}_{n-1}^{-1}, \quad u_{nn} = a_{nn} - \boldsymbol{\gamma}^{\mathrm{T}}\boldsymbol{\beta}$$

这就证明了 $\boldsymbol{A}$ 的 LU 分解的存在唯一性. □

矩阵的三角分解在求解线性方程组时将十分简便. 如对线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$，设 $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$ 是其 LU 分解. 我们先求解方程组 $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{b}$. 由于 $\boldsymbol{L}$ 是下三角矩阵，则解向量 $\boldsymbol{y}$ 可以通过依次求出其分量 $y_1, y_2, \cdots, y_n$ 而求出，再求解方程组 $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$. 解向量 $\boldsymbol{x}$ 可以通过该方程组依次求出分量 $x_n, x_{n-1}, \cdots, x_2, x_1$ 而快速得出. 于是由两个方程组 $\boldsymbol{U}\boldsymbol{x} = \boldsymbol{y}$, $\boldsymbol{L}\boldsymbol{y} = \boldsymbol{b}$ 的求解而给出 $\boldsymbol{L}\boldsymbol{U}\boldsymbol{x} = \boldsymbol{L}\boldsymbol{y} = \boldsymbol{b} = \boldsymbol{A}\boldsymbol{x}$ 的解.

当 $n$ 阶矩阵 $\boldsymbol{A}$ 满足定理 4.1.1 的条件时，我们可以用初等变换的方法求出 $\boldsymbol{L}$ 和 $\boldsymbol{U}$. 因为当 $\boldsymbol{A} = \boldsymbol{L}\boldsymbol{U}$ 时，由于 $\boldsymbol{L}$ 可逆，故必存在可逆阵 $\boldsymbol{P}$ 组

$$\boldsymbol{P}\boldsymbol{L} = \boldsymbol{E}$$

即

$$PA = PLU = U$$

也就是说,我们可以先对 $A$ 施行行的初等变换得出上三角阵 $U$,而矩阵 $P$ 可以通过对单位阵 $E$ 进行相同的行初等变换得出. 即有

$$P(A, E) = (PA, PE) = (U, P)$$

于是

$$A = P^{-1}U$$

为保持 $P$ 为下三角阵(从而 $P^{-1}$ 也是下三角阵),在进行行初等变换时,不能进行行的对换,上行的倍数应加到下行的对应元.

**例 4.1.1** 将矩阵

$$A = \begin{pmatrix} 1 & 0 & 7 \\ 3 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

进行 LU 分解.

**解**

$$(A, E) = \begin{pmatrix} 1 & 0 & 7 & 1 & 0 & 0 \\ 3 & 2 & 0 & 0 & 1 & 0 \\ 1 & 1 & 1 & 0 & 0 & 1 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 1 & 0 & 7 & 1 & 0 & 0 \\ 0 & 2 & -21 & -3 & 1 & 0 \\ 0 & 1 & -6 & -1 & 0 & 1 \end{pmatrix}$$

$$\rightarrow \begin{pmatrix} 1 & 0 & 7 & 1 & 0 & 0 \\ 0 & 2 & -21 & -3 & 1 & 0 \\ 0 & 0 & \frac{9}{2} & \frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix}$$

从而

$$P = \begin{pmatrix} 1 & 0 & 0 \\ -3 & 1 & 0 \\ \frac{1}{2} & -\frac{1}{2} & 1 \end{pmatrix}$$

$$U = \begin{pmatrix} 1 & 0 & 7 \\ 0 & 2 & -21 \\ 0 & 0 & \frac{9}{2} \end{pmatrix}$$

$$PA = U$$

因为

$$P^{-1} = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 1 & \frac{1}{2} & 1 \end{pmatrix}$$

所以

$$A = \begin{pmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 1 & \frac{1}{2} & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 7 \\ 0 & 2 & -21 \\ 0 & 0 & \frac{9}{2} \end{pmatrix} = LU$$

**定理 4.1.2** 在定理 4.1.1 的条件下，存在唯一的主对角元都是 1 的下三角阵 $L$ 与上三角阵 $U$，以及唯一的对角阵 $D$，使得

$$A = LDU$$

**证** 我们首先证明这种形式的分解的存在性. 由于

$$A = LU$$

其中 $L$ 是主对角元为 1 的下三角阵. 作分块矩阵的乘法：

$$A = \begin{pmatrix} A_k & A_{12} \\ A_{21} & A_{22} \end{pmatrix} = \begin{pmatrix} L_k & 0 \\ L_{21} & L_{22} \end{pmatrix} \begin{pmatrix} U_k & U_{12} \\ 0 & U_{22} \end{pmatrix}$$

则

$$A_k = L_k U_k$$

于是

$$|A_k| = |L_k||U_k| = u_{11}u_{22}\cdots u_{kk} \neq 0, \quad k = 1, 2, \cdots, n-1$$

即

$$A=\begin{pmatrix}1&0&\cdots&0\\l_{21}&1&\cdots&0\\\vdots&\vdots&&\vdots\\l_{n1}&l_{n2}&\cdots&1\end{pmatrix}\begin{pmatrix}u_{11}&u_{12}&\cdots&u_{1n}\\0&u_{22}&\cdots&u_{2n}\\\vdots&\vdots&&\vdots\\0&0&\cdots&u_{nn}\end{pmatrix}=LU$$

时，有 $u_{11}, u_{22}, \cdots, u_{n-1,n-1}\neq 0$. 所以

$$A=\begin{pmatrix}1&0&\cdots&0\\l_{21}&1&\cdots&0\\\vdots&\vdots&&\vdots\\l_{n1}&l_{n2}&\cdots&1\end{pmatrix}\begin{pmatrix}u_{11}&&&&\\&u_{22}&&&\\&&\ddots&&\\&&&u_{n-1,n-1}&\\&&&&u_{nn}\end{pmatrix}\begin{pmatrix}1&\frac{u_{12}}{u_{11}}&\frac{u_{13}}{u_{11}}&\cdots&\frac{u_{1,n-1}}{u_{11}}&\frac{u_{1n}}{u_{11}}\\0&1&\frac{u_{23}}{u_{22}}&\cdots&\frac{u_{2,n-1}}{u_{22}}&\frac{u_{2n}}{u_{22}}\\\vdots&\vdots&\vdots&&\vdots&\vdots\\0&0&0&\cdots&1&\frac{u_{n-1,n}}{u_{n-1,n-1}}\\0&0&0&\cdots&0&1\end{pmatrix}$$

$$=LDU_1$$

$L, U_1$ 就是主对角元全为 1 的下、上三角矩阵，$D$ 是对角阵. □

假定 $A=L_1D_1U_1=L_2D_2U_2$，其中 $L_1, L_2$ 是主对角元为 1 的下三角阵，$U_1, U_2$ 是主对角元为 1 的上三角阵，则

$$A=L_1(D_1U_1)=L_2(D_2U_2) \tag{4.1.1}$$

是 $A$ 的两个 $LU$ 分解. 根据 $A$ 的 $LU$ 分解的唯一性知

$$L_1=L_2=L,\qquad D_1U_1=D_2U_2=U$$

但由式(4.1.1)又得

$$L_2^{-1}L_1D_1=D_2U_2U_1^{-1} \tag{4.1.2}$$

式(4.1.2)左边的矩阵是下三角矩阵，且主对角线上的元素是 $D_1$ 的主对角线上的对应元，而右边的矩阵是上三角矩阵，其主对角线上的元素是 $D_2$ 的主对角线上的对应元. 从而 $L_2^{-1}L_1D_1=D_2U_2U_1^{-1}$ 是对角阵. 所以，$D_1=D_2$.

令 $D_1=D_2=\mathrm{diag}(d_1, d_2, \cdots, d_n)$，$U_1=(u_{ij}^{(1)})$，$U_2=(u_{ij}^{(2)})$，$U=(u_{ij})$，则由

$$D_1U_1=D_2U_2=U$$

得

$$
\begin{pmatrix}
d_1 & d_1u_{12}^{(1)} & d_1u_{13}^{(1)} & \cdots & d_1u_{1,n-1}^{(1)} & d_1u_{1n}^{(1)} \\
0 & d_2 & d_2u_{23}^{(1)} & \cdots & d_2u_{2,n-1}^{(1)} & d_2u_{2n}^{(1)} \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & \cdots & d_{n-1} & d_{n-1}u_{n-1,n}^{(1)} \\
0 & 0 & 0 & \cdots & 0 & d_n
\end{pmatrix}
$$

$$
=\begin{pmatrix}
d_1 & d_1u_{12}^{(2)} & d_1u_{13}^{(2)} & \cdots & d_1u_{1,n-1}^{(2)} & d_1u_{1n}^{(2)} \\
0 & d_2 & d_2u_{23}^{(2)} & \cdots & d_2u_{2,n-1}^{(2)} & d_2u_{2n}^{(2)} \\
\vdots & \vdots & \vdots & & \vdots & \vdots \\
0 & 0 & 0 & \cdots & d_{n-1} & d_{n-1}u_{n-1,n}^{(2)} \\
0 & 0 & 0 & \cdots & 0 & d_n
\end{pmatrix}
$$

$$
=\begin{pmatrix}
u_{11} & u_{12} & \cdots & u_{1n} \\
0 & u_{22} & \cdots & u_{2n} \\
\vdots & \vdots & & \vdots \\
0 & 0 & \cdots & u_{nn}
\end{pmatrix}
$$

由于已有 $u_{11}u_{22}\cdots u_{n-1,n-1}\neq 0$,故 $d_1=u_{11}$, $d_2=u_{22}$, $\cdots$, $d_{n-1}=u_{n-1}$,均不为零. 所以,有

$$d_iu_{ij}^{(1)}=d_iu_{ij}^{(2)},\quad i=1,2,\cdots,n-1;\ j=2,3,\cdots,n,\ i<j$$

即

$$u_{ij}^{(1)}=u_{ij}^{(2)},\ 1\leqslant i\leqslant n-1;\ 2\leqslant j\leqslant n,\ i<j$$

所以

$$\boldsymbol{U}_1=\boldsymbol{U}_2$$

**推论 4.1.1** 假定矩阵 $\boldsymbol{A}$ 满足定理 4.1.1 的条件且 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$,则存在主对角元为 1 的下三角阵 $\boldsymbol{L}$ 与实对角阵 $\boldsymbol{D}$,使得

$$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{D}\boldsymbol{L}^{\mathrm{H}}$$

称为 $\boldsymbol{A}$ 的 Cholesky(乔列斯基)分解.

**证** 由定理 4.1.1,存在主对角元为 1 的下、上三角阵 $\boldsymbol{L}$, $\boldsymbol{U}$ 及对角阵 $\boldsymbol{D}$,使

$$\boldsymbol{A}=\boldsymbol{L}\boldsymbol{D}\boldsymbol{U}$$

并且这样的表示是唯一的. 根据

$$\boldsymbol{A}^{\mathrm{H}}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{D}^{\mathrm{H}}\boldsymbol{L}^{\mathrm{H}}=\boldsymbol{L}\boldsymbol{D}\boldsymbol{U}=\boldsymbol{A}$$

及 $U^{\mathrm{H}}$, $L^{\mathrm{H}}$ 分别是主对角元为 1 的下、上三角阵，由分解的唯一性，可知

$$U = L^{\mathrm{H}}, \qquad D^{\mathrm{H}} = D$$

所以，$D$ 是实对角阵且

$$A = LDL^{\mathrm{H}}$$

□

## 4.2 矩阵的 QR 分解

在定理 2.2.1 中，我们利用 Gram-Schmidt(格拉姆-施密特)方法将一组线性无关的向量变为与之等价的一组单位正交向量. 这一过程将给我们提供另一种矩阵的分解形式.

**定理 4.2.1** 设 $A \in \mathbf{C}^{m\times n}$, $m \geqslant n$ 且 $\operatorname{rank} A = n$，则必存在非奇异上三角 $n\times n$ 矩阵 $R$ 及 $m\times n$ 矩阵 $Q$, $Q^{\mathrm{H}}Q = E_n$，使得

$$A = QR$$

**证** 将 $A$ 写成分块矩阵

$$A = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)$$

则因 $\operatorname{rank} A = n$，故 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 线性无关，由推论 2.2.1 知

$$(\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n) = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n)\begin{pmatrix} \|\boldsymbol{\beta}_1\| & (\boldsymbol{\alpha}_2, \boldsymbol{\gamma}_1) & \cdots & (\boldsymbol{\alpha}_n, \boldsymbol{\gamma}_1) \\ & \|\boldsymbol{\beta}_2\| & \cdots & (\boldsymbol{\alpha}_n, \boldsymbol{\gamma}_2) \\ & & \ddots & \vdots \\ & & & \|\boldsymbol{\beta}_n\| \end{pmatrix}$$

令

$$Q = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_n)$$

$$R = \begin{pmatrix} \|\boldsymbol{\beta}_1\| & (\boldsymbol{\alpha}_2, \boldsymbol{\gamma}_1) & \cdots & (\boldsymbol{\alpha}_n, \boldsymbol{\gamma}_1) \\ & \|\boldsymbol{\beta}_2\| & \cdots & (\boldsymbol{\alpha}_n, \boldsymbol{\gamma}_2) \\ & & \ddots & \vdots \\ & & & \|\boldsymbol{\beta}_n\| \end{pmatrix}$$

则因 $Q$ 的列向量组是单位正交组，故 $Q^{\mathrm{H}}Q = E_n$，又因 $\|\boldsymbol{\beta}_i\| \neq 0$ $(i = 1, 2, \cdots, n)$，故 $R$ 的主对角线上的元都不为 0，从而是非奇异的上三角阵.

特别地，当 $m=n$ 时，$\boldsymbol{Q}$ 是 $n$ 阶酉矩阵.

我们将等式 $\boldsymbol{A}=\boldsymbol{QR}$ 称为矩阵的 QR 分解. □

**例 4.2.1** 已知矩阵

$$\boldsymbol{A}=\begin{pmatrix}1&2&2\\1&0&2\\0&1&1\end{pmatrix}$$

求 $\boldsymbol{A}$ 的 QR 分解.

**解** 令 $\boldsymbol{A}=(\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3)$，对 $\boldsymbol{\alpha}_1,\boldsymbol{\alpha}_2,\boldsymbol{\alpha}_3$ 进行 Gram-Schmidt 过程得

$$\boldsymbol{\beta}_1=\boldsymbol{\alpha}_1=(1,1,0)^{\mathrm{T}},\qquad \boldsymbol{\gamma}_1=\frac{\boldsymbol{\beta}_1}{\|\boldsymbol{\beta}_1\|}=\frac{1}{\sqrt{2}}(1,1,0)^{\mathrm{T}}$$

$$\boldsymbol{\beta}_2=\boldsymbol{\alpha}_2-(\boldsymbol{\alpha}_2,\boldsymbol{\gamma}_1)\boldsymbol{\gamma}_1=(1,-1,1)^{\mathrm{T}},\qquad \boldsymbol{\gamma}_2=\frac{\boldsymbol{\beta}_2}{\|\boldsymbol{\beta}_2\|}=\frac{1}{\sqrt{3}}(1,-1,1)^{\mathrm{T}}$$

$$\boldsymbol{\beta}_3=\boldsymbol{\alpha}_3-(\boldsymbol{\alpha}_3,\boldsymbol{\gamma}_1)\boldsymbol{\gamma}_1-(\boldsymbol{\alpha}_3,\boldsymbol{\gamma}_2)\boldsymbol{\gamma}_2=\frac{1}{3}(-1,1,2)^{\mathrm{T}},\quad \boldsymbol{\gamma}_3=\frac{\boldsymbol{\beta}_3}{\|\boldsymbol{\beta}_3\|}=\frac{1}{\sqrt{6}}(-1,1,2)^{\mathrm{T}}$$

从而

$$\boldsymbol{A}=(\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\boldsymbol{\gamma}_3)\begin{pmatrix}\|\boldsymbol{\beta}_1\| & (\boldsymbol{\alpha}_2,\boldsymbol{\gamma}_1) & (\boldsymbol{\alpha}_3,\boldsymbol{\gamma}_1)\\ & \|\boldsymbol{\beta}_2\| & (\boldsymbol{\alpha}_3,\boldsymbol{\gamma}_2)\\ & & \|\boldsymbol{\beta}_3\|\end{pmatrix}$$

$$=\begin{pmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}}\\ 0 & \frac{1}{\sqrt{3}} & \frac{2}{\sqrt{6}}\end{pmatrix}\begin{pmatrix}\sqrt{2} & \sqrt{2} & 2\sqrt{2}\\ & \sqrt{3} & \frac{1}{\sqrt{3}}\\ & & \frac{\sqrt{6}}{3}\end{pmatrix}=\boldsymbol{QR}$$

即为所求的 QR 分解.

对于矩阵是列满秩时，利用 Gram-Schmidt 方法可以得到矩阵的 QR 分解；如果矩阵不是列满秩时，其实也可以利用这一方法得到矩阵的 QR 分解.

**例 4.2.2** 已知矩阵

$$\boldsymbol{A}=\begin{pmatrix}1&2&1\\0&0&1\\1&2&1\end{pmatrix}$$

求 $A$ 的 QR 分解.

**解** 令 $A=(\alpha_1, \alpha_2, \alpha_3)$，对 $\alpha_1, \alpha_2, \alpha_3$ 进行 Gram-Schmidt 过程得

$$\beta_1=\alpha_1=(1, 0, 1)^T, \qquad \gamma_1=\frac{\beta_1}{\|\beta_1\|}=\frac{1}{\sqrt{2}}(1, 0, 1)^T$$

$$\beta_2=\alpha_2-(\alpha_2, \gamma_1)\gamma_1=(0, 0, 0)^T, \qquad \gamma_2=(0, 0, 0)^T$$

$$\beta_3=\alpha_3-(\alpha_3, \gamma_1)\gamma_1-(\alpha_3, \gamma_2)\gamma_2=(0, 1, 0)^T, \quad \gamma_3=\frac{\beta_3}{\|\beta_3\|}=(0, 1, 0)^T$$

从而

$$A=(\gamma_1, \gamma_2, \gamma_3)\begin{pmatrix}\|\beta_1\| & (\alpha_2, \gamma_1) & (\alpha_3, \gamma_1)\\ & \|\beta_2\| & (\alpha_3, \gamma_2)\\ & & \|\beta_3\|\end{pmatrix}$$

$$=\begin{pmatrix}\frac{1}{\sqrt{2}} & 0 & 0\\ 0 & 0 & 1\\ \frac{1}{\sqrt{2}} & 0 & 0\end{pmatrix}\begin{pmatrix}\sqrt{2} & \sqrt{2} & \sqrt{2}\\ 0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}$$

由于上面第二式左边矩阵不是正交阵，所以第二式不是矩阵 $A$ 的 $QR$ 分解. 但是上面第二式中左边矩阵中的第二列取任何值都不影响等式，故可取第二列为 $\left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)^T$，于是

$$A=\begin{pmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0\\ 0 & 0 & 1\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0\end{pmatrix}\begin{pmatrix}\sqrt{2} & \sqrt{2} & \sqrt{2}\\ 0 & 0 & 0\\ 0 & 0 & 1\end{pmatrix}$$

即为 $A$ 的 QR 分解.

对于不是列满秩的矩阵，也可以利用所谓的 Householder 方法得到矩阵的 QR 分解.

**定义 4.2.1** 设 $u\in \mathbf{C}^n$ 是单位向量，令

$$H=E-2uu^H$$

称 $H$ 为 Householder 矩阵或 Householder 变换.

**定理 4.2.2** 设 $\boldsymbol{H}$ 是 $n$ 阶 Householder 矩阵,则

(i) $\boldsymbol{H}^{\mathrm{H}}=\boldsymbol{H}$;

(ii) $\boldsymbol{H}^2=\boldsymbol{E}$;

(iii) $\boldsymbol{H}$ 以 1 为 $n-1$ 重特征值,$-1$ 为单特征值;

(iv) $\widetilde{\boldsymbol{H}}=\begin{pmatrix}1 & \boldsymbol{0}^{\mathrm{T}}\\ \boldsymbol{0} & \boldsymbol{H}\end{pmatrix}$是 $n+1$ 阶 Householder 矩阵,其中 $\boldsymbol{0}\in\mathbf{C}^n$.

**证** 容易由定义证明(i)和(ii).下证(iii)与(iv).

(iii) 当 $n=1$ 时,$\boldsymbol{H}=(-1)$ 结论成立.假定 $n\geqslant 2$. 若 $\lambda$ 为 $\boldsymbol{H}$ 的一个特征值,$\boldsymbol{\alpha}$ 为对应于 $\lambda$ 的特征向量,则

$$\boldsymbol{H\alpha}=\lambda\boldsymbol{\alpha}$$

上式左乘 $\boldsymbol{H}$,由(ii)得

$$(\lambda^2-1)\boldsymbol{\alpha}=\boldsymbol{0}$$

因 $\boldsymbol{\alpha}\neq\boldsymbol{0}$,故

$$\lambda=\pm 1$$

设特征值 1 与 $-1$ 的重数分别为 $r_1$,$r_2$,则

$$r_1+r_2=n$$

令 $\boldsymbol{u}=(\xi_1,\xi_2,\cdots,\xi_n)^{\mathrm{T}}$ 是单位向量,使得 $\boldsymbol{H}=\boldsymbol{E}-2\boldsymbol{uu}^{\mathrm{H}}$,则

$$\begin{aligned}\mathrm{Tr}(\boldsymbol{H})&=(1-2\mid\xi_1\mid^2)+(1-2\mid\xi_2\mid^2)+\cdots+(1-2\mid\xi_n\mid^2)\\&=n-2\sum_{i=1}^{n}\mid\xi_i\mid^2=n-2\boldsymbol{u}^{\mathrm{H}}\boldsymbol{u}=n-2=r_1-r_2\end{aligned}$$

于是 $r_1$,$r_2$ 满足方程组

$$\begin{cases}r_1+r_2=n\\ r_1-r_2=n-2\end{cases}$$

由该方程组解得

$$r_1=n-1,\qquad r_2=1$$

(iv) $$\begin{aligned}\widetilde{\boldsymbol{H}}&=\begin{pmatrix}1 & \boldsymbol{0}^{\mathrm{T}}\\ \boldsymbol{0} & \boldsymbol{H}\end{pmatrix}=\begin{pmatrix}1 & \boldsymbol{0}^{\mathrm{T}}\\ \boldsymbol{0} & \boldsymbol{E}_n-2\boldsymbol{uu}^{\mathrm{H}}\end{pmatrix}=\begin{pmatrix}1 & \boldsymbol{0}^{\mathrm{T}}\\ \boldsymbol{0} & \boldsymbol{E}_n\end{pmatrix}-\begin{pmatrix}0 & \boldsymbol{0}^{\mathrm{T}}\\ \boldsymbol{0} & -2\boldsymbol{uu}^{\mathrm{H}}\end{pmatrix}\\&=\boldsymbol{E}_{n+1}-2\begin{pmatrix}0\\ \boldsymbol{u}\end{pmatrix}(0,\boldsymbol{u}^{\mathrm{H}})\end{aligned}$$

令

$$\tilde{\boldsymbol{u}} = \begin{pmatrix} 0 \\ \boldsymbol{u} \end{pmatrix}$$

则 $\tilde{\boldsymbol{u}} \in \mathbf{C}^{n+1}$，且$\tilde{\boldsymbol{u}}^{\mathrm{H}} \tilde{\boldsymbol{u}} = (0, \boldsymbol{u}^{\mathrm{H}}) \begin{pmatrix} 0 \\ \boldsymbol{u} \end{pmatrix} = \boldsymbol{u}^{\mathrm{H}} \boldsymbol{u} = 1$，故$\tilde{\boldsymbol{u}}$ 是 $n+1$ 维单位向量. 所以

$$\widetilde{\boldsymbol{H}} = \boldsymbol{E}_{n+1} - 2\, \tilde{\boldsymbol{u}}\, \tilde{\boldsymbol{u}}^{\mathrm{H}}$$

是 $n+1$ 阶 Householder 矩阵. □

由定理 4.2.2 立即可得：$|\boldsymbol{H}| = -1$，$\begin{pmatrix} \boldsymbol{E}_r & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H} \end{pmatrix}$ 仍为 Householder 矩阵.

对矩阵进行 QR 分解时，主要利用 Householder 矩阵的下列结果.

**定理 4.2.3** 设 $\boldsymbol{e}_1 = (1, 0, \cdots, 0)^{\mathrm{T}} \in \mathbf{C}^n$，$\boldsymbol{\beta} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$，$\boldsymbol{\beta} \neq \boldsymbol{0}$. 令

$$\mu = \begin{cases} \dfrac{\|\boldsymbol{\beta}\|}{|x_1|} x_1, & x_1 \neq 0 \\ \|\boldsymbol{\beta}\|, & x_1 = 0 \end{cases}$$

$$\boldsymbol{u} = \frac{\boldsymbol{\beta} + \mu \boldsymbol{e}_1}{\|\boldsymbol{\beta} + \mu \boldsymbol{e}_1\|}$$

则 $\boldsymbol{H} = \boldsymbol{E} - 2\boldsymbol{u}\boldsymbol{u}^{\mathrm{H}}$ 是 $n$ 阶 Householder 矩阵，且有

$$\boldsymbol{H}\boldsymbol{\beta} = -\mu \boldsymbol{e}_1$$

**证** 首先注意到，因为

$$\boldsymbol{\beta} + \mu \boldsymbol{e}_1 = \begin{cases} \left(\left(1 + \dfrac{\|\boldsymbol{\beta}\|}{|x_1|}\right) x_1, x_2, \cdots, x_n\right)^{\mathrm{T}}, & x_1 \neq 0 \\ (\|\boldsymbol{\beta}\|, x_2, \cdots, x_n)^{\mathrm{T}}, & x_1 = 0 \end{cases}$$

所以

$$\boldsymbol{\beta} + \mu \boldsymbol{e}_1 \neq \boldsymbol{0}$$

从而 $\boldsymbol{u}$ 是有定义的单位向量. 又

$$\mu \boldsymbol{\beta}^{\mathrm{H}} \boldsymbol{e}_1 = \mu \bar{x}_1 = \begin{cases} |x_1| \, \|\boldsymbol{\beta}\|, & x_1 \neq 0 \\ 0, & x_1 = 0 \end{cases}$$

故 $\mu \boldsymbol{\beta}^{\mathrm{H}} \boldsymbol{e}_1$ 是实数，且 $|\mu| = \|\boldsymbol{\beta}\|$. 于是

$$\begin{aligned} \boldsymbol{H\beta} &= \boldsymbol{\beta} - 2\boldsymbol{uu}^{\mathrm{H}}\boldsymbol{\beta} \\ &= \boldsymbol{\beta} - (\boldsymbol{\beta} + \mu\boldsymbol{e}_1)\frac{2(\boldsymbol{\beta}+\mu\boldsymbol{e}_1)^{\mathrm{H}}\boldsymbol{\beta}}{\|\boldsymbol{\beta}+\mu\boldsymbol{e}_1\|^2} \end{aligned}$$

由于

$$\begin{aligned} (\boldsymbol{\beta}+\mu\boldsymbol{e}_1)^{\mathrm{H}}(\boldsymbol{\beta}+\mu\boldsymbol{e}_1) &= (\boldsymbol{\beta}^{\mathrm{H}}+\bar{\mu}\boldsymbol{e}_1^{\mathrm{H}})(\boldsymbol{\beta}+\mu\boldsymbol{e}_1) \\ &= \boldsymbol{\beta}^{\mathrm{H}}\boldsymbol{\beta}+\bar{\mu}\boldsymbol{e}_1^{\mathrm{H}}\boldsymbol{\beta}+\mu\boldsymbol{\beta}^{\mathrm{H}}\boldsymbol{e}_1+|\mu|^2 \\ &= 2\boldsymbol{\beta}^{\mathrm{H}}\boldsymbol{\beta}+2\bar{\mu}\boldsymbol{e}_1^{\mathrm{H}}\boldsymbol{\beta} \\ &= 2(\boldsymbol{\beta}+\mu\boldsymbol{e}_1)^{\mathrm{H}}\boldsymbol{\beta} \end{aligned}$$

所以

$$\boldsymbol{H\beta} = \boldsymbol{\beta} - (\boldsymbol{\beta}+\mu\boldsymbol{e}_1) = -\mu\boldsymbol{e}_1$$

上述定理反映了 Householder 变换的作用，即可以将非零向量 $\boldsymbol{\alpha}$ 变成与 $\boldsymbol{e}_1$ 共线的向量 $-\sigma\boldsymbol{e}_1$，而且 $|\sigma| = \|\boldsymbol{\alpha}\|$.

**定理 4.2.4** 设 $\boldsymbol{A}$ 为任一 $n$ 阶矩阵，则必存在 $n$ 阶酉矩阵 $\boldsymbol{Q}$ 和 $n$ 阶上三角阵 $\boldsymbol{R}$，使得

$$\boldsymbol{A} = \boldsymbol{QR}$$

**证** 第一步，将矩阵 $\boldsymbol{A}$ 按列分块写成 $\boldsymbol{A} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)$. 如果 $\boldsymbol{\alpha}_1 \neq \boldsymbol{0}$，则由定理 4.2.4 知，存在 $n$ 阶 Householder 矩阵 $\boldsymbol{H}_1$，使得

$$\boldsymbol{H}_1\boldsymbol{\alpha}_1 = -a_1\boldsymbol{e}_1, \qquad |a_1| = \|\boldsymbol{\alpha}_1\|, \quad \boldsymbol{e}_1 \in \mathbf{C}^n$$

于是有

$$\boldsymbol{H}_1\boldsymbol{A} = (\boldsymbol{H}_1\boldsymbol{\alpha}_1, \boldsymbol{H}_1\boldsymbol{\alpha}_2, \cdots, \boldsymbol{H}_1\boldsymbol{\alpha}_n) = \left(\begin{array}{c|ccc} -a_1 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & \boldsymbol{A}_{n-1} & \\ 0 & & & \end{array}\right)$$

如果 $\boldsymbol{\alpha}_1 = \boldsymbol{0}$，则直接进行下一步，此时相当于取 $\boldsymbol{H}_1 = \boldsymbol{E}_n$，而 $a_1 = 0$.

第二步，将 $\boldsymbol{A}_{n-1}$ 按列分块写成 $\boldsymbol{A}_{n-1} = (\boldsymbol{\beta}_2, \boldsymbol{\beta}_3, \cdots, \boldsymbol{\beta}_n)$，如果 $\boldsymbol{\beta}_2 \neq \boldsymbol{0}$，则存在 $n-1$ 阶 Householder 矩阵 $\widetilde{\boldsymbol{H}}_2$，使得

$$\widetilde{\boldsymbol{H}}_2\boldsymbol{\beta}_2 = -a_2\tilde{\boldsymbol{e}}_1, \qquad |a_2| = \|\boldsymbol{\beta}_1\|, \quad \tilde{\boldsymbol{e}}_1 \in \mathbf{C}^{n-1}$$

于是有

$$\widetilde{\boldsymbol{H}}_2\boldsymbol{A}_{n-1} = (\widetilde{\boldsymbol{H}}_2\boldsymbol{\beta}_2, \widetilde{\boldsymbol{H}}_2\boldsymbol{\beta}_3, \cdots, \widetilde{\boldsymbol{H}}_2\boldsymbol{\beta}_n)$$

$$
= \left(\begin{array}{c|ccc} -a_2 & * & \cdots & * \\ \hline 0 & & & \\ \vdots & & A_{n-2} & \\ 0 & & & \end{array}\right)
$$

此时,令

$$
H_2 = \begin{pmatrix} 1 & \mathbf{0}^{\mathrm{T}} \\ \mathbf{0} & \widetilde{H}_2 \end{pmatrix}
$$

则 $H_2$ 是 $n$ 阶 Householder 矩阵,且使

$$
H_2 H_1 A = \left(\begin{array}{cc|ccc} -a_1 & * & * & \cdots & * \\ 0 & -a_2 & * & \cdots & * \\ \hline 0 & & & A_{n-2} & \end{array}\right)
$$

当 $\beta_2 = \mathbf{0}$ 时,直接进行下一步.

第三步,对 $n-2$ 阶矩阵继续进行类似的变换,如此下去,至多在第 $n-1$ 步,我们可以找到 Householder 矩阵 $H_1, H_2, \cdots, H_{n-1}$,使得

$$
H_{n-1} \cdots H_2 H_1 A = \begin{pmatrix} -a_1 & & & & * \\ & -a_2 & & & \\ & & \ddots & & \\ & & & -a_{n-1} & \\ & & & & a_n \end{pmatrix} = R
$$

其中可能有某些

$$
H_j = E_n
$$

令 $Q^{-1} = H_{n-1} \cdots H_2 H_1$,则 $Q$ 是酉矩阵之积,从而必为酉矩阵并且

$$
A = QR
$$

□

**例 4.2.3** 对矩阵

$$
A = \begin{pmatrix} 1 & 1 & 0 \\ 2 & 1 & 1 \\ 2 & 0 & 2 \end{pmatrix}
$$

利用 Householder 方法求其 QR 分解.

**解** 对向量 $\alpha_1 = (1, 2, 2)^{\mathrm{T}}$,取 $a_1 = -3$,令

$$u_1 = \frac{\alpha_1 + a_1 e_1}{\| \alpha_1 + a_1 e_1 \|} = \frac{1}{\sqrt{3}}(-1, 1, 1)^{\mathrm{T}}$$

由此作 Householder 矩阵

$$H_1 = E - 2u_1 u_1^{\mathrm{H}} = \frac{1}{3}\begin{pmatrix} 1 & 2 & 2 \\ 2 & 1 & -2 \\ 2 & -2 & 1 \end{pmatrix}$$

从而

$$H_1 A = \begin{pmatrix} 3 & 1 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} = R$$

所以矩阵 $A$ 的 QR 分解为

$$A = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & -\frac{2}{3} & \frac{1}{3} \end{pmatrix}\begin{pmatrix} 3 & 1 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \end{pmatrix} = QR$$

## 4.3 矩阵的满秩分解

非零矩阵的满秩分解是一种形式最为简单的分解类型. 我们将在本节对此进行讨论. 本章的后面将会看到其应用.

**定理 4.3.1** 设 $A$ 是 $m \times n$ 复矩阵，$\operatorname{rank} A = r \geqslant 1$，则必存在秩为 $r$ 的两个矩阵 $P_1 \in \mathbf{C}^{m \times r}$，$Q_1 \in \mathbf{C}^{r \times n}$，使得

$$A = P_1 Q_1 \tag{4.3.1}$$

**证** 在线性代数中，我们已经知道，对任意秩为 $r$ 的矩阵 $A$，必可以通过初等变换将其化成 Hermite 标准形. 即有 $m$ 阶满秩矩阵 $P$ 与 $n$ 阶满秩矩阵 $Q$，使得

$$P^{-1} A Q^{-1} = \begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}$$

将 $P$，$Q$ 分块

$$P = (P_1, P_2), \qquad Q = \begin{pmatrix} Q_1 \\ Q_2 \end{pmatrix}$$

其中 $P_1 \in \mathbf{C}^{m\times r}$，$Q_1 \in \mathbf{C}^{r\times n}$，则 $P_1$ 为列满秩矩阵，$Q_1$ 为行满秩矩阵. 并且有

$$A = P\begin{pmatrix} I_r & 0 \\ 0 & 0 \end{pmatrix}Q = P\left[\begin{pmatrix} I_r \\ 0 \end{pmatrix}(I_r, 0)\right]Q = \left[P\begin{pmatrix} I_r \\ 0 \end{pmatrix}\right][(I_r, 0)Q] = P_1Q_1 \qquad \square$$

我们将式(4.3.1)称为 $A$ 的满秩分解. 定理的证明过程虽然也提供了一种矩阵的满秩分解方法，但计算比较麻烦. 下面来讨论其简便的做法.

在线性代数中，我们知道行最简形矩阵是一种特殊形式的矩阵，这种矩阵具有如下特征：

(1) 元素全为 0 的行在非零行的下方.

(2) 非零行中从左到右的第一个非零元是 1，称为该行的首 1 元素. 首 1 元是所在列的唯一非零元.

任一矩阵必可以通过有限次行的初等变换化成行最简形.

**定理 4.3.2** 设 $A$ 是 $m\times n$ 矩阵，$\operatorname{rank} A = r$. 如果 $A$ 的行最简形矩阵 $H$ 的首 1 元分别在 $H$ 的第 $j_1, j_2, \cdots, j_r$ 列，取 $H$ 的前 $r$ 行构成的矩阵为 $C$，$A$ 的第 $j_1, j_2, \cdots, j_r$ 列构成的矩阵为 $B$，则

$$A = BC \tag{4.3.2}$$

就是矩阵 $A$ 的一个满秩分解.

**证** 由于 $A$ 经过行的初等变换化成 $H$，则必存在 $m$ 阶可逆矩阵 $P$，使得 $PA = H$. 将 $A$，$H$ 按列分块

$$A = (\alpha_1, \alpha_2, \cdots, \alpha_n)$$

$$H = (P\alpha_1, P\alpha_2, \cdots, P\alpha_n) = (\tilde{\alpha}_1, \tilde{\alpha}_2, \cdots, \tilde{\alpha}_n)$$

由行最简形矩阵的定义不难看出

$$P\alpha_{j_i} = \tilde{\alpha}_{j_i} = e_{j_i} = (0, \cdots, 0, \overset{(j_i)}{1}, 0, \cdots, 0)^{\mathrm{T}} \in \mathbf{C}^m$$

其中，$i = 1, 2, \cdots, r$；$\tilde{\alpha}_{j_1}, \tilde{\alpha}_{j_2}, \cdots, \tilde{\alpha}_{j_r}$ 是 $H$ 的列向量组的极大线性无关组.

将矩阵 $C$ 按列分块

$$C = (\beta_1, \beta_2, \cdots, \beta_n), \quad \beta_k \in \mathbf{C}^r; \ k = 1, 2, \cdots, n$$

易知 $C \in \mathbf{C}^{r\times n}$ 且 $\operatorname{rank} C = r$，且 $\beta_1, \cdots, \beta_n$ 是向量 $\tilde{\alpha}_1, \cdots, \tilde{\alpha}_n$ 分别由极大线性无关组 $\tilde{\alpha}_{j1}, \tilde{\alpha}_{j2}, \cdots, \tilde{\alpha}_{jr}$ 线性表示的向量，即

$$(\widetilde{\boldsymbol{\alpha}}_1, \widetilde{\boldsymbol{\alpha}}_2, \cdots, \widetilde{\boldsymbol{\alpha}}_n) = (\widetilde{\boldsymbol{\alpha}}_{j1}, \widetilde{\boldsymbol{\alpha}}_{j2}, \cdots, \widetilde{\boldsymbol{\alpha}}_{jr})(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n) \tag{4.3.3}$$

令 $\boldsymbol{B} = (\widetilde{\boldsymbol{\alpha}}_{j1}, \widetilde{\boldsymbol{\alpha}}_{j2}, \cdots, \widetilde{\boldsymbol{\alpha}}_{jr}) \in \mathbf{C}^{m\times r}$，则由(4.3.3)有

$$\begin{aligned}
\boldsymbol{A} &= \boldsymbol{P}^{-1}(\widetilde{\boldsymbol{\alpha}}_1, \widetilde{\boldsymbol{\alpha}}_2, \cdots, \widetilde{\boldsymbol{\alpha}}_n) \\
&= (\boldsymbol{P}^{-1}\widetilde{\boldsymbol{\alpha}}_{j1}, \boldsymbol{P}^{-1}\widetilde{\boldsymbol{\alpha}}_{j2}, \cdots, \boldsymbol{P}^{-1}\widetilde{\boldsymbol{\alpha}}_{jr})(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n) \\
&= (\widetilde{\boldsymbol{\alpha}}_{j1}, \widetilde{\boldsymbol{\alpha}}_{j2}, \cdots, \widetilde{\boldsymbol{\alpha}}_{jr})(\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_n) \\
&= \boldsymbol{BC}
\end{aligned}$$

□

**例 4.3.1** 设矩阵

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 2 & 1 \end{pmatrix}$$

由于矩阵为列满秩，所以其满秩分解为

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \\ 2 & 1 \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

**例 4.3.2** 设矩阵

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 3 & 7 \\ 1 & 0 & 6 & 5 \\ 0 & 2 & 6 & -4 \end{pmatrix}$$

试求 $\boldsymbol{A}$ 的满秩分解.

**解** 对 $\boldsymbol{A}$ 施行行初等变换

$$\boldsymbol{A} = \begin{pmatrix} 1 & -1 & 3 & 7 \\ 1 & 0 & 6 & 5 \\ 0 & 2 & 6 & -4 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 6 & 5 \\ 0 & 1 & 3 & -2 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

取

$$\boldsymbol{B} = \begin{pmatrix} 1 & -1 \\ 1 & 0 \\ 0 & 2 \end{pmatrix}, \qquad \boldsymbol{C} = \begin{pmatrix} 1 & 0 & 6 & 5 \\ 0 & 1 & 3 & -2 \end{pmatrix}$$

则 $\boldsymbol{A} = \boldsymbol{BC}$ 即为 $\boldsymbol{A}$ 的满秩分解.

## 4.4 矩阵的奇异值分解

由第 2 章知，对于正规矩阵 $\boldsymbol{A} \in \mathbf{C}^{n \times n}$，存在酉矩阵 $\boldsymbol{U}$ 及对角阵 $\boldsymbol{D}$，使得

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\boldsymbol{U}^{\mathrm{H}}$$

对于一般的 $m \times n$ 矩阵，我们希望找到类似的分解，这就是奇异值分解. 矩阵的奇异值分解是计算矩阵的重要手段，在控制理论、优化问题、广义逆矩阵等有着重要的应用.

由第 2 章知，对于 $m \times n$ 矩阵 $\boldsymbol{A}$，由于 $n$ 阶矩阵 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是正规矩阵，从而必酉相似于一个对角阵，且其特征值都是非负实数

$$\lambda_1 \geqslant \cdots \geqslant \lambda_r > 0 = \lambda_{r+1} \geqslant \cdots \geqslant \lambda_n$$

其中，$\operatorname{rank}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}) = \operatorname{rank} \boldsymbol{A} = r$. 称 $d_i = \sqrt{\lambda_i}$ $(i = 1, \cdots, n)$为矩阵 $\boldsymbol{A}$ 的奇异值. 显然，奇异值是半正定(或正定)矩阵$(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{1/2}$的特征值，其中非零奇异值的个数等于矩阵 $\boldsymbol{A}$ 的秩.

**定理 4.4.1** 设矩阵 $\boldsymbol{A} \in \mathbf{C}^{m \times n}$，$\operatorname{rank} \boldsymbol{A} = r$，则存在酉矩阵 $\boldsymbol{U} \in \mathbf{C}^{m \times m}$，$\boldsymbol{V} \in \mathbf{C}^{n \times n}$，使得

$$\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^{\mathrm{H}} \tag{4.4.1}$$

其中，$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{D} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$；$\boldsymbol{D} = \operatorname{diag}(d_1, \cdots, d_r)$ $(d_1 \geqslant \cdots \geqslant d_r)$ 为矩阵 $\boldsymbol{A}$ 的正奇异值. 式(4.4.1)称为矩阵 $\boldsymbol{A}$ 的奇异值分解.

**证** 首先，设矩阵 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的特征值为 $\lambda_1 \geqslant \cdots \geqslant \lambda_r > 0 = \lambda_{r+1} \geqslant \cdots \geqslant \lambda_n$ 和相对应的正交单位特征向量 $\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_r, \boldsymbol{\gamma}_{r+1}, \cdots, \boldsymbol{\gamma}_n$. 令

$$\boldsymbol{D} = \operatorname{diag}(d_1, \cdots, d_r), \qquad d_i = \sqrt{\lambda_i}, \quad i = 1, \cdots, r$$

$$\boldsymbol{V} = (\boldsymbol{V}_1, \boldsymbol{V}_2), \quad \boldsymbol{V}_1 = (\boldsymbol{\gamma}_1, \cdots, \boldsymbol{\gamma}_r), \quad \boldsymbol{V}_2 = (\boldsymbol{\gamma}_{r+1}, \cdots, \boldsymbol{\gamma}_n)$$

于是矩阵 $\boldsymbol{V}$ 为酉矩阵，且

$$\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{V} = \begin{pmatrix} \boldsymbol{D}^2 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{pmatrix}$$

亦即

$$\begin{bmatrix} V_1^H \\ V_2^H \end{bmatrix} A^H A(V_1, V_2) = \begin{bmatrix} V_1^H A^H A V_1 & V_1^H A^H A V_2 \\ V_2^H A^H A V_1 & V_2^H A^H A V_2 \end{bmatrix} = \begin{bmatrix} D^2 & 0 \\ 0 & 0 \end{bmatrix}$$

比较上面等式,有

$$V_1^H A^H A V_1 = (AV_1)^H (AV_1) = D^2$$

$$V_2^H A^H A V_2 = (AV_2)^H (AV_2) = 0$$

由此有

$$D^{-1} V_1^H A^H A V_1 D^{-1} = E_r \tag{4.4.2}$$

$$AV_2 = 0 \tag{4.4.3}$$

令$U_1 = AV_1 D^{-1} \in \mathbf{C}^{m\times r}$,则由式(4.4.2)知$U_1$的列向量是$r$个$m$维的正交的单位向量,故一定存在矩阵$U_2 \in \mathbf{C}^{m\times(m-r)}$,使得矩阵$U = (U_1, U_2)$为酉矩阵. 于是

$$U_2^H U_1 = 0 \tag{4.4.4}$$

最后由式(4.4.2)~式(4.4.4)有

$$U^H A V = \begin{bmatrix} U_1^H \\ U_2^H \end{bmatrix} A(V_1, V_2) = \begin{bmatrix} U_1^H A V_1 D^{-1} D & U_1^H A V_2 \\ U_2^H A V_1 D^{-1} D & U_2^H A V_2 \end{bmatrix}$$

$$= \begin{bmatrix} U_1^H U_1 D & U_1^H A V_2 \\ U_2^H U_1 D & U_2^H A V_2 \end{bmatrix} = \begin{pmatrix} D & 0 \\ 0 & 0 \end{pmatrix} = \Sigma$$

具体计算矩阵的奇异值分解时,可以按下列步骤进行:

(1) 求出矩阵$A^H A$的特征值为$\lambda_1 \geqslant \cdots \geqslant \lambda_r > 0 = \lambda_{r+1} \geqslant \cdots \geqslant \lambda_n$,确定$A$的非零奇异值$d_i = \sqrt{\lambda_i}$ $(i = 1, \cdots, r)$.

(2) 求出矩阵$A^H A$的对应于特征值为$\lambda_i$的特征向量并将其单位正交化,得到标准正交向量组$\gamma_1, \cdots, \gamma_r, \gamma_{r+1}, \cdots, \gamma_n$. 令

$$V = (V_1, V_2), \quad V_1 = (\gamma_1, \cdots, \gamma_r), \quad V_2 = (\gamma_{r+1}, \cdots, \gamma_n)$$

(3) 计算$U_1 = AV_1 D^{-1}$. 取一矩阵$U_2 \in \mathbf{C}^{m\times(m-r)}$,使得它的$m-r$个列向量和$U_1$的$r$个列向量构成$\mathbf{C}^m$的标准正交基. 令$U = (U_1, U_2)$,从而$A = U\Sigma V^H$.

**例 4.4.1** 求矩阵$A = \begin{bmatrix} 1 & 2 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$的奇异值分解.

**解** 因

$$A^HA=\begin{pmatrix}1&2\\2&4\end{pmatrix}$$

故 $A$ 的奇异值为

$$d_1=\sqrt{5},\qquad d_2=0$$

$A^HA$ 对应于特征值 5,0 的标准正交特征向量为

$$\gamma_1=\left[\frac{1}{\sqrt{5}},\frac{2}{\sqrt{5}}\right]^T,\qquad \gamma_2=\left[\frac{2}{\sqrt{5}},\frac{-1}{\sqrt{5}}\right]^T$$

令

$$V=\begin{bmatrix}\frac{1}{\sqrt{5}}&\frac{2}{\sqrt{5}}\\\frac{2}{\sqrt{5}}&\frac{-1}{\sqrt{5}}\end{bmatrix}=(V_1,V_2),\qquad V_1=(\gamma_1),\quad V_2=(\gamma_2)$$

计算 $U_1=AV_1D^{-1}=\begin{bmatrix}1\\0\\0\end{bmatrix}$，取 $U_2=\begin{bmatrix}0&0\\1&0\\0&1\end{bmatrix}$，则

$$U=(U_1,U_2)=E_3$$

从而

$$U^HAV=\begin{bmatrix}\sqrt{5}&0\\0&0\\0&0\end{bmatrix}$$

## 4.5 广义逆矩阵

作为矩阵分解的一个理论上的应用，我们在本节将讨论广义逆矩阵的定义与构造方法. 由广义逆矩阵可以确定线性方程组的最佳逼近解.

**定义 4.5.1** 设 $A\in C^{m\times n}$，若有矩阵 $X\in C^{n\times m}$，使得

$$AXA=A \tag{4.5.1}$$

$$XAX=X \tag{4.5.2}$$

$$(AX)^H=AX \tag{4.5.3}$$

$$(\boldsymbol{XA})^{\mathrm{H}} = \boldsymbol{XA} \tag{4.5.4}$$

则称 $\boldsymbol{X}$ 为矩阵 $\boldsymbol{A}$ 的 Moore-Penrose(摩尔-彭罗斯)逆,简称为 $\boldsymbol{A}$ 的加号逆,记为 $\boldsymbol{A}^{+}$.

当 $\boldsymbol{A}$ 是 $n$ 阶可逆阵时,容易看出 $\boldsymbol{A}^{+} = \boldsymbol{A}^{-1}$.

**定理 4.5.1** 对任意的 $\boldsymbol{A} \in \mathbf{C}^{m\times n}$, $\boldsymbol{A}^{+}$存在且唯一.

**证** 先证明 $\boldsymbol{A}^{+}$ 的存在性.

若 $\boldsymbol{A} = \boldsymbol{0}$,取 $\boldsymbol{X} = \boldsymbol{0} \in \mathbf{C}^{n\times m}$, 则显然

$$\boldsymbol{X} = \boldsymbol{A}^{+}$$

不妨设 $\operatorname{rank} \boldsymbol{A} = r \geqslant 1$,并设 $\boldsymbol{A} = \boldsymbol{BC}$ 为 $\boldsymbol{A}$ 的满秩分解,其中 $\boldsymbol{B} \in \mathbf{C}^{m\times r}$, $\boldsymbol{C} \in \mathbf{C}^{r\times n}$,且 $\operatorname{rank} \boldsymbol{B} = \operatorname{rank} \boldsymbol{C} = r$. 令

$$\boldsymbol{X} = \boldsymbol{C}^{\mathrm{H}}(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}} \tag{4.5.5}$$

我们证明 $\boldsymbol{X}$ 就是 $\boldsymbol{A}$ 的加号逆.

由

$$\operatorname{rank}(\boldsymbol{CC}^{\mathrm{H}}) = r,\ \operatorname{rank}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}) = r$$

得 $\boldsymbol{CC}^{\mathrm{H}}$, $\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}$ 是 $r$ 阶满秩阵,$\boldsymbol{X}$ 是有定义的.

$$\begin{aligned}
\boldsymbol{AXA} &= \boldsymbol{BC}[\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}]\boldsymbol{BC} \\
&= \boldsymbol{B}(\boldsymbol{CC}^{\mathrm{H}})(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})\boldsymbol{C} \\
&= \boldsymbol{BC} = \boldsymbol{A}
\end{aligned}$$

$$\begin{aligned}
\boldsymbol{XAX} &= [\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}]\boldsymbol{BC}[\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}] \\
&= \boldsymbol{C}^{\mathrm{H}}(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})(\boldsymbol{CC}^{\mathrm{H}})(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}} \\
&= \boldsymbol{C}^{\mathrm{H}}(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}} \\
&= \boldsymbol{X}
\end{aligned}$$

$$\begin{aligned}
(\boldsymbol{AX})^{\mathrm{H}} &= [\boldsymbol{BCC}^{\mathrm{H}}(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}]^{\mathrm{H}} \\
&= [\boldsymbol{B}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}]^{\mathrm{H}} \\
&= \boldsymbol{B}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}} \\
&= \boldsymbol{B}(\boldsymbol{CC}^{\mathrm{H}})(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}} \\
&= \boldsymbol{BC}[\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}] \\
&= \boldsymbol{AX}
\end{aligned}$$

$$(\boldsymbol{XA})^{\mathrm{H}} = [\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{CC}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{BC}]^{\mathrm{H}}$$

$$
\begin{aligned}
&= [\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}\boldsymbol{C}]^{\mathrm{H}} \\
&= \boldsymbol{C}^{\mathrm{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}\boldsymbol{C} \\
&= \boldsymbol{C}^{\mathrm{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})\boldsymbol{C} \\
&= \boldsymbol{X}\boldsymbol{A}
\end{aligned}
$$

$\boldsymbol{X}$ 满足式(4.5.1)～式(4.5.4)，从而

$$\boldsymbol{X} = \boldsymbol{A}^{+}$$

再证唯一性.

假定 $\boldsymbol{X}_1$，$\boldsymbol{X}_2$ 均为 $\boldsymbol{A}$ 的加号逆，则

$$
\begin{aligned}
\boldsymbol{X}_1 &\overset{(2)}{=} \boldsymbol{X}_1\boldsymbol{A}\boldsymbol{X}_1 \overset{(4)}{=} (\boldsymbol{X}_1\boldsymbol{A})^{\mathrm{H}}\boldsymbol{X}_1 = \boldsymbol{A}^{\mathrm{H}}\boldsymbol{X}_1^{\mathrm{H}}\boldsymbol{X}_1 \overset{(1)}{=} (\boldsymbol{A}\boldsymbol{X}_2\boldsymbol{A})^{\mathrm{H}}\boldsymbol{X}_1^{\mathrm{H}}\boldsymbol{X}_1 = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{X}_2^{\mathrm{H}})\boldsymbol{A}^{\mathrm{H}}\boldsymbol{X}_1^{\mathrm{H}}\boldsymbol{X}_1 \\
&\overset{(4)}{=} (\boldsymbol{X}_2\boldsymbol{A})(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{X}_1^{\mathrm{H}})\boldsymbol{X}_1 \overset{(4)}{=} (\boldsymbol{X}_2\boldsymbol{A})(\boldsymbol{X}_1\boldsymbol{A})\boldsymbol{X}_1 \overset{(3)}{=} \boldsymbol{X}_2\boldsymbol{A}\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}} \\
&\overset{(1)}{=} \boldsymbol{X}_2\boldsymbol{A}\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{X}_2\boldsymbol{A})^{\mathrm{H}} = \boldsymbol{X}_2\boldsymbol{A}\boldsymbol{X}_1\boldsymbol{X}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{X}_2)^{\mathrm{H}} \\
&\overset{(3)}{=} \boldsymbol{X}_2\boldsymbol{A}\boldsymbol{X}_1(\boldsymbol{X}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}})\boldsymbol{A}\boldsymbol{X}_2 \overset{(3)}{=} \boldsymbol{X}_2(\boldsymbol{A}\boldsymbol{X}_1\boldsymbol{A})\boldsymbol{X}_1\boldsymbol{A}\boldsymbol{X}_2 \\
&\overset{(1)}{=} \boldsymbol{X}_2\boldsymbol{A}\boldsymbol{X}_1\boldsymbol{A}\boldsymbol{X}_2 \overset{(1)}{=} \boldsymbol{X}_2\boldsymbol{A}\boldsymbol{X}_2 \overset{(2)}{=} \boldsymbol{X}_2
\end{aligned}
$$

□

**推论 4.5.1** 当 $\boldsymbol{A}$ 是行满秩矩阵时

$$\boldsymbol{A}^{+} = \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1} \tag{4.5.6}$$

当 $\boldsymbol{A}$ 是列满秩矩阵时

$$\boldsymbol{A}^{+} = (\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{H}} \tag{4.5.7}$$

**证** 设 $\boldsymbol{A} \in \mathbf{C}^{m\times n}$，$\operatorname{rank} \boldsymbol{A} = m$，$\boldsymbol{A} = \boldsymbol{B}\boldsymbol{C}$ 为 $\boldsymbol{A}$ 的秩分解，则 $\boldsymbol{B}$ 是 $m$ 阶可逆矩阵，于是

$$
\begin{aligned}
\boldsymbol{A}^{+} &= \boldsymbol{C}^{\mathrm{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}} \\
&= \boldsymbol{C}^{\mathrm{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}\boldsymbol{B}^{-1} \\
&= \boldsymbol{C}^{\mathrm{H}}\boldsymbol{B}^{\mathrm{H}}[(\boldsymbol{B}^{\mathrm{H}})^{-1}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}\boldsymbol{B}^{-1}] \\
&= \boldsymbol{A}^{\mathrm{H}}[\boldsymbol{B}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})\boldsymbol{B}^{\mathrm{H}}]^{-1} \\
&= \boldsymbol{A}^{\mathrm{H}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{-1}
\end{aligned}
$$

类似可以证列满秩的情况. □

以下定理列出了 $\boldsymbol{A}^{+}$ 的一些主要性质.

**定理 4.5.2** 设 $\boldsymbol{A}^{+}$ 是矩阵 $\boldsymbol{A}$ 的 Moore-Penrose 逆，则

(i) $(\boldsymbol{A}^{+})^{+} = \boldsymbol{A}$；

(ii) $(\lambda \boldsymbol{A})^{+}=\lambda^{-1}\boldsymbol{A}^{+}$, $\lambda\neq 0$;

(iii) $(\boldsymbol{A}^{k})^{+}=(\boldsymbol{A}^{+})^{k}$, $k$ 是正整数,$\boldsymbol{A}$ 是正规矩阵;

(iv) $(\boldsymbol{A}^{\mathrm{H}})^{+}=(\boldsymbol{A}^{+})^{\mathrm{H}}$;

(v) $(\boldsymbol{A}^{\mathrm{T}})^{+}=(\boldsymbol{A}^{+})^{\mathrm{T}}$;

(vi) $(\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}})^{+}=(\boldsymbol{A}^{+})^{\mathrm{H}}\boldsymbol{A}^{+}$, $(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A})^{+}=\boldsymbol{A}^{+}(\boldsymbol{A}^{+})^{\mathrm{H}}$;

(vii) $(\boldsymbol{U}\boldsymbol{A}\boldsymbol{V})^{+}=\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}^{+}\boldsymbol{U}^{\mathrm{H}}$, $\boldsymbol{U}$, $\boldsymbol{V}$ 是酉矩阵.

**证** 我们只选证(v)和(vii). 其他等式要么是显然的,或者可以类似地证明.

(v)的证明. 若 $\boldsymbol{A}=\boldsymbol{B}\boldsymbol{C}$ 是 $\boldsymbol{A}$ 的秩分解,则 $\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{C}^{\mathrm{T}}\boldsymbol{B}^{\mathrm{T}}$ 就是 $\boldsymbol{A}^{\mathrm{T}}$ 的秩分解. 于是

$$\begin{aligned}(\boldsymbol{A}^{\mathrm{T}})^{+}&=(\boldsymbol{B}^{\mathrm{T}})^{\mathrm{H}}[\boldsymbol{B}^{\mathrm{T}}(\boldsymbol{B}^{\mathrm{T}})^{\mathrm{H}}]^{-1}[(\boldsymbol{C}^{\mathrm{T}})^{\mathrm{H}}\boldsymbol{C}^{\mathrm{T}}]^{-1}(\boldsymbol{C}^{\mathrm{T}})^{\mathrm{H}}\\&=(\boldsymbol{B}^{\mathrm{H}})^{\mathrm{T}}[(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{\mathrm{T}}]^{-1}[(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{\mathrm{T}}]^{-1}(\boldsymbol{C}^{\mathrm{H}})^{\mathrm{T}}\\&=[(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}]^{\mathrm{T}}(\boldsymbol{C}^{\mathrm{H}})^{\mathrm{T}}\\&=[\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}]^{\mathrm{T}}\\&=(\boldsymbol{A}^{+})^{\mathrm{T}}\end{aligned}$$

(vii)的证明. 因 $\boldsymbol{U}\boldsymbol{A}\boldsymbol{V}=(\boldsymbol{U}\boldsymbol{B})(\boldsymbol{C}\boldsymbol{V})$ 是 $\boldsymbol{U}\boldsymbol{A}\boldsymbol{V}$ 的秩分解,故

$$\begin{aligned}(\boldsymbol{U}\boldsymbol{A}\boldsymbol{V})^{+}&=(\boldsymbol{C}\boldsymbol{V})^{\mathrm{H}}[\boldsymbol{C}\boldsymbol{V}(\boldsymbol{C}\boldsymbol{V})^{\mathrm{H}}]^{-1}[(\boldsymbol{U}\boldsymbol{B})^{\mathrm{H}}(\boldsymbol{U}\boldsymbol{B})]^{-1}(\boldsymbol{U}\boldsymbol{B})^{\mathrm{H}}\\&=\boldsymbol{V}^{\mathrm{H}}[\boldsymbol{C}^{\mathrm{H}}(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}]\boldsymbol{U}^{\mathrm{H}}\\&=\boldsymbol{V}^{\mathrm{H}}\boldsymbol{A}^{+}\boldsymbol{U}^{\mathrm{H}}\end{aligned}$$

□

**例 4.5.1** 设矩阵

$$\boldsymbol{A}=\begin{pmatrix}1&2\\1&2\\1&2\end{pmatrix}$$

试求 $\boldsymbol{A}^{+}$.

**解** 先求 $\boldsymbol{A}$ 的满秩分解为 $\boldsymbol{A}^{+}$.

$$\boldsymbol{A}=\begin{pmatrix}1\\1\\1\end{pmatrix}(1,\ 2)=\boldsymbol{B}\boldsymbol{C}$$

由此

$$\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}}=(5),\qquad(\boldsymbol{C}\boldsymbol{C}^{\mathrm{H}})^{-1}=\left(\frac{1}{5}\right)$$

$$\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B}=(3),\qquad(\boldsymbol{B}^{\mathrm{H}}\boldsymbol{B})^{-1}=\left(\frac{1}{3}\right)$$

所以

$$A^{+}=C^{H}(CC^{H})^{-1}(B^{H}B)^{-1}B^{H}=\frac{1}{15}\begin{pmatrix}1&1&1\\2&2&2\end{pmatrix}$$

**例 4.5.2** 设 $D=(s_{ij})$ 是 $m\times n$ 矩阵,且当 $i\neq j$ 时, $s_{ij}=0$,称 $D$ 为广义对角阵.令 $X=(t_{ke})$ 为 $n\times m$ 矩阵,其中

$$t_{ii}=\begin{cases}s_{ii}^{-1}, & s_{ii}\neq 0\\ 0, & s_{ii}=0\end{cases}$$

$(i=1, 2, \cdots, \min(m, n))k\neq e$ 时, $t_{ke}=0$. 容易验证 $X$ 满足式(4.5.1)~式(4.5.4).所以, $X=D^{+}$.

下面的定理是利用矩阵的奇异值分解求广义逆.

**定理 4.5.3** 设 $A=UDV^{H}$ 是矩阵 $A$ 的奇异值分解,则

$$A^{+}=VD^{+}U^{H}$$

**证** 利用定理 4.5.2(iii)及例 4.5.2 即证. □

回顾定理 4.4.1 的证明过程,我们还可以将利用矩阵的奇异值分解求 $A^{+}$ 的方法进一步简化.

设矩阵 $A\in\mathbf{C}^{m\times n}$, $\operatorname{rank}A=r$, $A^{H}A$ 的特征值为

$$\lambda_1\geqslant\cdots\geqslant\lambda_r>0=\lambda_{r+1}\geqslant\cdots\geqslant\lambda_n$$

相对应的正交的单位特征向量 $\gamma_1, \cdots, \gamma_r, \gamma_{r+1}, \cdots, \gamma_n$.令

$$\Sigma=\begin{pmatrix}D&0\\0&0\end{pmatrix},\quad D=\operatorname{diag}(d_1, \cdots, d_r),\quad d_i=\sqrt{\lambda_i},\quad i=1, \cdots, r$$

则存在酉矩阵 $U$, $V$ 使得

$$A=U\Sigma V^{H}$$

其中

$$V=(V_1, V_2),\quad V_1=(\gamma_1, \cdots, \gamma_r)\in\mathbf{C}^{n\times r},\quad V_2=(\gamma_{r+1}, \cdots, \gamma_n)$$

$$U=(U_1, U_2),\quad U_1=AV_1D^{-1}\in\mathbf{C}^{m\times r}$$

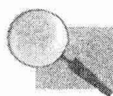
于是

$$A=(U_1, U_2)\begin{pmatrix}D&0\\0&0\end{pmatrix}\begin{bmatrix}V_1^{H}\\V_2^{H}\end{bmatrix}=U_1DV_1^{H}$$

注意，$U_1$，$V_1$ 并非酉矩阵，但 $\operatorname{rank} U_1 = \operatorname{rank}(DV_1^H) = \operatorname{rank}(V_1) = r$，且 $U_1^H U_1 = E_r$，$V_1^H V_1 = E_r$. 所以，$A = U_1(DV_1^H)$ 是 $A$ 的一个满秩分解，利用定理 4.5.1，有

$$\begin{aligned} A^+ &= (DV_1^H)^H[DV_1^H(DV_1^H)^H]^{-1}[U_1^H U_1]^{-1}U_1^H \\ &= V_1 D^{-1} U_1^H \\ &= V_1 D^{-1}(AV_1 D^{-1})^H \\ &= V_1(D^2)^{-1}V_1^H A^H \end{aligned}$$

这里 $(D^2)^{-1} = \operatorname{diag}(\lambda_1^{-1}, \lambda_2^{-1}, \cdots, \lambda_r^{-1})$.



## 习 题 4

1. 试求下列矩阵的 LU 分解：

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 4 & 1 & 0 \\ -2 & 2 & 1 \end{pmatrix}$$

2. 试求下列矩阵的 QR 分解：

(1) $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$；(2) $\begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 0 & 1 \end{pmatrix}$；(3) $\begin{pmatrix} 0 & 4 & 1 \\ 1 & 1 & 1 \\ 0 & 3 & 2 \end{pmatrix}$.

3. 试对下列矩阵作秩分解

$$A = \begin{pmatrix} 1 & 2 & 3 & 0 \\ 0 & 2 & 1 & -1 \\ -2 & 4 & -2 & -4 \end{pmatrix}$$

4. 设 $B, C \in \mathbf{R}^{n\times n}$，$A = B + \mathrm{i}C$，且

$$R = \begin{pmatrix} B & -C \\ C & B \end{pmatrix}, \qquad W = \begin{pmatrix} B & C \\ C & B \end{pmatrix}$$

证明：

(1) 若 $A$ 是正规矩阵，则 $R$ 是实正规矩阵；

(2) 若 $A$ 是 Hermite 矩阵，则 $R$ 是对称矩阵；

(3) 若 $A$ 是正定矩阵，则 $W$ 是正定矩阵；

(4) 若 $\boldsymbol{A}$ 是酉矩阵，则 $\boldsymbol{R}$ 是正交矩阵.

5. 设 $\boldsymbol{A}$ 是 $n$ 阶正规矩阵，证明 $\boldsymbol{A}$ 的奇异值为

$$s_i = |\lambda_i|, \quad i=1, \cdots, n$$

其中 $\lambda_i$ 是 $\boldsymbol{A}$ 的特征值.

6. 设 $\boldsymbol{A}$ 是 $n$ 阶正规矩阵，证明：

$$|\det \boldsymbol{A}| = s_1 s_2 \cdots s_n$$

其中 $s_i$ 是 $\boldsymbol{A}$ 的奇异值.

7. 试求矩阵

$$\boldsymbol{A} = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix}$$

的奇异值分解.

8. 验证下列两个矩阵

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \qquad \boldsymbol{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

互为另一个矩阵的 Moore-Penrose 逆.

9. 设 $\boldsymbol{A} \in \mathbf{C}^{m\times r}$, $\boldsymbol{B} \in \mathbf{C}^{r\times n}$ $(r \leqslant \min(m, n))$ 分别为列满秩矩阵和行满秩矩阵. 证明：

$$(\boldsymbol{AB})^+ = \boldsymbol{B}^+ \boldsymbol{A}^+$$

10. 设

$$\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 2 & 0 \\ -1 & 3 \end{bmatrix}$$

求 $\boldsymbol{A}^+$.

# 第5章 范数理论及其应用

在第2章中我们用内积定义了向量的长度，这个向量长度是几何向量长度概念的一种推广. 本章采用公理化的方法把向量长度概念进一步推广，即范数的概念. 范数在数值计算和矩阵分析中起着重要的作用.

## 5.1 向量范数

**定义 5.1.1** 设 $V$ 是数域 $P$ 上的线性空间. 若对于 $V$ 中任一向量 $\boldsymbol{\alpha}$，都有一非负实数 $\|\boldsymbol{\alpha}\|$ 与之对应，并且满足下列三个条件：

(i) 正定性. 当 $\boldsymbol{\alpha} \neq \mathbf{0}$ 时，都有

$$\|\boldsymbol{\alpha}\| > 0$$

(ii) 齐次性. 对任何 $k \in P$，都有

$$\|k\boldsymbol{\alpha}\| = |k|\ \|\boldsymbol{\alpha}\|$$

(iii) 三角不等式. 对任何向量 $\boldsymbol{\alpha}, \boldsymbol{\beta} \in V$，都有

$$\|\boldsymbol{\alpha}+\boldsymbol{\beta}\| \leqslant \|\boldsymbol{\alpha}\| + \|\boldsymbol{\beta}\|$$

则称 $\|\boldsymbol{\alpha}\|$ 为向量 $\boldsymbol{\alpha}$ 的范数. 定义了范数的线性空间称为赋范线性空间.

**例 5.1.1** 设 $\boldsymbol{\alpha} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{R}^n$，$\boldsymbol{\alpha}$ 的长度 $\|\boldsymbol{\alpha}\| = \sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}$ 就是 $\mathbf{R}^n$ 上的一种范数.

事实上，当 $\boldsymbol{\alpha} \neq \mathbf{0}$ 时，至少有一分量不为0，所以 $\|\boldsymbol{\alpha}\| > 0$；当 $\boldsymbol{\alpha} = \mathbf{0}$ 时，$\|\boldsymbol{\alpha}\| = \sqrt{0^2 + 0^2 + \cdots + 0^2} = 0$.

又对任意实数 $k \in \mathbf{R}$，因为 $k\boldsymbol{\alpha} = (kx_1, kx_2, \cdots, kx_n)$，所以

$$\begin{aligned}\|k\boldsymbol{\alpha}\| &= \sqrt{(kx_1)^2 + \cdots + (kx_n)^2} \\ &= |k|\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}\end{aligned}$$

$$=|k|\ \|\boldsymbol{\alpha}\|$$

最后证明三角不等式. 对 $\mathbf{R}^n$ 中任意的两个向量 $\boldsymbol{\alpha}$ 与 $\boldsymbol{\beta}$,由柯西-施瓦茨不等式可得

$$\begin{aligned}\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|^2 &= \|\boldsymbol{\alpha}\|^2+2(\boldsymbol{\alpha},\boldsymbol{\beta})+\|\boldsymbol{\beta}\|^2 \\ &\leqslant \|\boldsymbol{\alpha}\|^2+2\|\boldsymbol{\alpha}\|\ \|\boldsymbol{\beta}\|+\|\boldsymbol{\beta}\|^2 \\ &= (\|\boldsymbol{\alpha}\|+\|\boldsymbol{\beta}\|)^2\end{aligned}$$

从而有

$$\|\boldsymbol{\alpha}+\boldsymbol{\beta}\| \leqslant \|\boldsymbol{\alpha}\|+\|\boldsymbol{\beta}\|$$

这就证明了 $\|\boldsymbol{\alpha}\|=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}$ 是 $\mathbf{R}^n$ 上的一种范数. 这种范数称为 2-范数或欧氏范数. 记为

$$\|\boldsymbol{\alpha}\|_2=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}=\left(\sum_{i=1}^n x_i^2\right)^{\frac{1}{2}} \tag{5.1.1}$$

**例 5.1.2**　设 $\boldsymbol{\alpha}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}\in\mathbf{C}^n$, 定义

$$\|\boldsymbol{\alpha}\|_1=\sum_{i=1}^n |x_i|$$

则 $\|\boldsymbol{\alpha}\|_1$ 是向量 $\boldsymbol{\alpha}$ 的一种范数,称为 1-范数.

**证**　(1) 正定性:

当 $\boldsymbol{\alpha}\neq\mathbf{0}$ 时

$$\|\boldsymbol{\alpha}\|_1=\sum_{i=1}^n |x_i|>0$$

当 $\boldsymbol{\alpha}=\mathbf{0}$ 时,$\boldsymbol{\alpha}$ 的每一个分量都是零,故

$$\|\boldsymbol{\alpha}\|_1=0$$

(2) 齐次性: $\forall k\in\mathbf{C}$, 有

$$\|k\boldsymbol{\alpha}\|_1=\sum_{i=1}^n |kx_i|=|k|\sum_{i=1}^n |x_i|=|k|\ \|\boldsymbol{\alpha}\|_1$$

(3) 三角不等式: $\forall\beta\in\mathbf{C}^n$, $\beta=(y_1, y_2, \cdots, y_n)^{\mathrm{T}}$, 有

$$\begin{aligned}\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_1 &= \sum_{i=1}^n |x_i+y_i| \leqslant \sum_{i=1}^n (|x_i|+|y_i|) \\ &= \sum_{i=1}^n |x_i|+\sum_{i=1}^n |y_i|\end{aligned}$$

$$= \|\boldsymbol{\alpha}\|_1 + \|\boldsymbol{\beta}\|_1$$

故 $\|\boldsymbol{\alpha}\|_1$ 是 $\mathbf{C}^n$ 上的一种范数.

**例 5.1.3** 设 $\boldsymbol{\alpha} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$, 定义

$$\|\boldsymbol{\alpha}\|_2 = \sqrt{\sum_{i=1}^{n} |x_i|^2} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{\frac{1}{2}}$$

则 $\|\boldsymbol{\alpha}\|_2$ 是向量 $\boldsymbol{\alpha}$ 的一种范数,称为 2-范数.

事实上,由例 5.1.1 即知.

**例 5.1.4** 设 $\boldsymbol{\alpha} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$, 定义

$$\|\boldsymbol{\alpha}\|_\infty = \max_i |x_i|$$

则 $\|\boldsymbol{\alpha}\|_\infty$ 是向量 $\boldsymbol{\alpha}$ 的一种范数.

**证** $\boldsymbol{\alpha}$ 的正定性、齐次性显然. 下证三角不等式:

设 $\boldsymbol{\beta} = (y_1, y_2, \cdots, y_n)^{\mathrm{T}} \in \mathbf{C}^n$, 有

$$\begin{aligned} \|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_\infty &= \max_i |x_i + y_i| \\ &\leqslant \max_i |x_i| + \max_i |y_i| \\ &= \|\boldsymbol{\alpha}\|_\infty + \|\boldsymbol{\beta}\|_\infty \end{aligned}$$

故 $\|\boldsymbol{\alpha}\|_\infty$ 是 $\mathbf{C}^n$ 上的一种范数.

**例 5.1.5** 设 $\boldsymbol{\alpha} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$, 定义

$$\|\boldsymbol{\alpha}\|_p = \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}}, \ 1 \leqslant p < +\infty$$

则 $\|\boldsymbol{\alpha}\|_p$ 是向量 $\boldsymbol{\alpha}$ 的一种范数,称为 $p$-范数.(证明略)

显然,当 $p = 1$ 时,即为例 5.1.2 的范数 $\|\boldsymbol{\alpha}\|_1$;当 $p = 2$ 时,即得到例 5.1.3 中的范数 $\|\boldsymbol{\alpha}\|_2$,当 $p = +\infty$ 时是否为$\infty$-范数呢? 我们有如下定理.

**定理 5.1.1** 设 $\boldsymbol{\alpha} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$, 则

$$\lim_{p \to +\infty} \|\boldsymbol{\alpha}\|_p = \|\boldsymbol{\alpha}\|_\infty \tag{5.1.2}$$

**证** 若 $\boldsymbol{\alpha} = \mathbf{0}$, 结论显然成立.

以下设 $\boldsymbol{\alpha} \neq \mathbf{0}$,又设 $|x_j| = \max_i |x_i| = \|\boldsymbol{\alpha}\|_\infty$, 则有

$$\begin{aligned} \|\boldsymbol{\alpha}\|_\infty = |x_j| &\leqslant \left(\sum_{i=1}^{n} |x_i|^p\right)^{\frac{1}{p}} \\ &= \|\boldsymbol{\alpha}\|_p \leqslant (n |x_j|^p)^{\frac{1}{p}} \end{aligned}$$

$$= n^{\frac{1}{p}} \| \boldsymbol{\alpha} \|_{\infty}$$

注意到 $\lim\limits_{p \to +\infty} n^{\frac{1}{p}} = 1$，由极限的夹逼准则有

$$\lim_{p \to +\infty} \| \boldsymbol{\alpha} \|_p = \| \boldsymbol{\alpha} \|_{\infty}$$ □

**例 5.1.6** 计算 $\mathbf{C}^4$ 中向量$=(3\mathrm{i}, 0, -4\mathrm{i}, -12)^{\mathrm{T}}$ 的 1，2，∞范数，这里 $\mathrm{i} = \sqrt{-1}$.

**解**
$$\| \boldsymbol{\alpha} \|_1 = \sum_{i=1}^{4} | x_i | = | 3\mathrm{i} | + | -4\mathrm{i} | + | -12 | = 19$$

$$\| \boldsymbol{\alpha} \|_2 = \left( \sum_{i=1}^{4} | x_i |^2 \right)^{\frac{1}{2}} = \sqrt{| 3\mathrm{i} |^2 + | -4\mathrm{i} |^2 + | -12 |^2} = 13$$

$$\| \boldsymbol{\alpha} \|_{\infty} = \max_i | x_i | = \max(3, 0, 4, 12) = 12$$

由此可见，在同一个线性空间中，不同定义的范数其大小可能不同.

**例 5.1.7** 在 $\mathbf{C}^n$ 中，$\boldsymbol{\alpha} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$，定义

$$\| \boldsymbol{\alpha} \|_p = \left( \sum_{i=1}^{n} | x_i |^p \right)^{\frac{1}{p}}, \quad 0 < p < 1$$

则 $\| \boldsymbol{\alpha} \|_p$ 不是 $\mathbf{C}^n$ 中的范数.

事实上，$\| \boldsymbol{\alpha} \|_p$ 不满足定义 5.1.1 中的条件(iii). 例如，取 $\boldsymbol{\alpha} = (1, 0, \cdots, 0)^{\mathrm{T}}$，$\boldsymbol{\beta} = (0, 1, 0, \cdots, 0)^{\mathrm{T}}$，$p = \dfrac{1}{2}$，则

$$\| \boldsymbol{\alpha} + \boldsymbol{\beta} \|_{\frac{1}{2}} = 4, \quad \| \boldsymbol{\alpha} \|_{\frac{1}{2}} = 1, \quad \| \boldsymbol{\beta} \|_{\frac{1}{2}} = 1$$

所以

$$\| \boldsymbol{\alpha} + \boldsymbol{\beta} \|_{\frac{1}{2}} \leqslant \| \boldsymbol{\alpha} \|_{\frac{1}{2}} + \| \boldsymbol{\beta} \|_{\frac{1}{2}}$$

不成立. 故 $\| \boldsymbol{\alpha} \|_{\frac{1}{2}}$ 不是 $\mathbf{C}^n$ 中的向量范数.

下面的例子，给出由已知的某范数去构造出新的向量范数的一种方法.

**例 5.1.8** 设 $\| \cdot \|_{\alpha}$ 是 $\mathbf{C}^m$ 上的一种向量范数，给定矩阵 $\boldsymbol{A} \in \mathbf{C}^{m \times n}$，且 $\boldsymbol{A}$ 的 $n$ 个列向量线性无关，对任意的 $\boldsymbol{\gamma} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$，规定

$$\| \boldsymbol{\gamma} \|_{\beta} = \| \boldsymbol{A} \boldsymbol{\gamma} \|_{\alpha}$$

则 $\| \boldsymbol{\gamma} \|_{\beta}$ 是 $\mathbf{C}^n$ 中的向量范数.

**证** (1) 设 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 是矩阵 $\boldsymbol{A}$ 的 $n$ 个线性无关的列向量，从而对任何 $\boldsymbol{\gamma} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \neq \mathbf{0}$，有

$$A\boldsymbol{\gamma} = (\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n)\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

$$= x_1\boldsymbol{\alpha}_1 + x_2\boldsymbol{\alpha}_2 + \cdots + x_n\boldsymbol{\alpha}_n \neq \mathbf{0}$$

由于$\|\cdot\|_\alpha$是$\mathbf{C}^m$上的向量范数,故

$$\|A\boldsymbol{\gamma}\|_\alpha > 0$$

即
$$\|\boldsymbol{\gamma}\|_\beta = \|A\boldsymbol{\gamma}\|_\alpha > 0$$

(2) 设 $\forall k \in \mathbf{C}$, $\forall \boldsymbol{\gamma} \in \mathbf{C}^n$, 则

$$\|k\boldsymbol{\gamma}\|_\beta = \|A(k\boldsymbol{\gamma})\|_\alpha = \|k(A\boldsymbol{\gamma})\|_\alpha$$
$$= |k|\ \|A\boldsymbol{\gamma}\|_\alpha = |k|\ \|\boldsymbol{\gamma}\|_\beta$$

(3) $\forall \boldsymbol{\gamma}, \boldsymbol{\eta} \in \mathbf{C}^n$, 有

$$\|\boldsymbol{\gamma}+\boldsymbol{\eta}\|_\beta = \|A(\boldsymbol{\gamma}+\boldsymbol{\eta})\|_\alpha = \|A\boldsymbol{\gamma}+A\boldsymbol{\eta}\|_\alpha$$
$$\leqslant \|A\boldsymbol{\gamma}\|_\alpha + \|A\boldsymbol{\eta}\|_\alpha$$
$$= \|\boldsymbol{\gamma}\|_\beta + \|\boldsymbol{\eta}\|_\beta$$

故$\|\boldsymbol{\gamma}\|_\beta$是$\mathbf{C}^n$中的向量范数.

以上我们一直在$\mathbf{C}^m$或$\mathbf{C}^n$上讨论向量范数,实际上向量的范数不限于此.

**例 5.1.9** 设 $V$ 是复数域 $\mathbf{C}$ 上的 $n$ 维线性空间, $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 是 $V$ 的一组基. 任意 $\boldsymbol{\alpha} \in V$ 可以唯一地表示为

$$\boldsymbol{\alpha} = x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n$$

记 $\boldsymbol{X} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}} \in \mathbf{C}^n$, 又设$\|\cdot\|$是$\mathbf{C}^n$上的向量范数,规定

$$\|\boldsymbol{\alpha}\|_\mu = \|\boldsymbol{X}\|$$

则$\|\boldsymbol{\alpha}\|_\mu$是 $V$ 中向量的一种范数.

**证** (1) 任意 $\boldsymbol{\alpha} \in V$, 若 $\boldsymbol{\alpha}$ 为非零向量,则其坐标 $\boldsymbol{X} \neq \mathbf{0}$, 从而

$$\|\boldsymbol{\alpha}\|_\mu = \|\boldsymbol{X}\| > 0$$

若 $\boldsymbol{\alpha}$ 为零向量,则其坐标 $\boldsymbol{X} = \mathbf{0}$, 于是

$$\|\boldsymbol{\alpha}\|_\mu = \|\boldsymbol{X}\| = 0$$

(2) 对任意的 $\boldsymbol{\alpha}\in V$, $k\in\mathbf{C}$, $k\boldsymbol{\alpha}=\sum_{i=1}^{n}k\boldsymbol{e}_i$, 即 $k\boldsymbol{\alpha}$ 的坐标为$(kx_1, kx_2, \cdots, kx_n)^{\mathrm{T}}\in\mathbf{C}^n$, 于是有

$$\|k\boldsymbol{\alpha}\|_{\mu}=\|k\boldsymbol{X}\|=|k|\ \|\boldsymbol{X}\|=|k|\ \|\boldsymbol{\alpha}\|_{\mu}$$

(3) 对任意的 $\boldsymbol{\beta}\in V$, 则

$$\boldsymbol{\beta}=\sum_{i=1}^{n}y_i\boldsymbol{e}_i,\ \boldsymbol{Y}=(y_1, y_2, \cdots, y_n)^{\mathrm{T}}\in\mathbf{C}^n,\ \boldsymbol{\alpha}+\boldsymbol{\beta}=\sum_{i=1}^{n}(x_i+y_i)\boldsymbol{e}_i$$

于是

$$\begin{aligned}\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_{\mu}&=\|\boldsymbol{X}+\boldsymbol{Y}\|\leqslant\|\boldsymbol{X}\|+\|\boldsymbol{Y}\|\\&=\|\boldsymbol{\alpha}\|_{\mu}+\|\boldsymbol{\beta}\|_{\mu}\end{aligned}$$

故 $\|\boldsymbol{\alpha}\|_{\mu}$ 是 $V$ 上的向量范数.

这样,抽象空间的向量范数可以用 $\mathbf{C}^n$ 上的向量范数来表示.

从前述讨论可知,同一线性空间上向量的范数可以有无穷多种(只要 $p$ 在 1 与 $\infty$之间取值). 但这些范数之间有重要的关系.

**定义 5.1.2** 设 $\|\boldsymbol{\alpha}\|_a$, $\|\boldsymbol{\alpha}\|_b$ 是 $n$ 维线性空间 $V$ 上定义的任意两种范数,若存在两个与 $\boldsymbol{\alpha}$ 无关的正常数 $c_1$, $c_2$,使得

$$c_1\|\boldsymbol{\alpha}\|_b\leqslant\|\boldsymbol{\alpha}\|_a\leqslant c_2\|\boldsymbol{\alpha}\|_b \tag{5.1.3}$$

则称 $\|\boldsymbol{\alpha}\|_a$ 与 $\|\boldsymbol{\alpha}\|_b$ 是等价的.

**定理 5.1.2** 有限维线性空间上的不同范数是等价的.

**证** 设 $\boldsymbol{e}_1, \boldsymbol{e}_2, \cdots, \boldsymbol{e}_n$ 是 $V$ 的一组基,于是 $V$ 中任意向量 $\boldsymbol{\alpha}$ 可以表示为

$$\boldsymbol{\alpha}=x_1\boldsymbol{e}_1+x_2\boldsymbol{e}_2+\cdots+x_n\boldsymbol{e}_n$$

定义

$$\|\boldsymbol{\alpha}\|_2=\sqrt{x_1^2+x_2^2+\cdots+x_n^2}$$

显然 $\|\boldsymbol{\alpha}\|_2$ 是一种范数. 若存在正常数 $c_1'$, $c_2'$ 和 $c_1''$, $c_2''$,使

$$c_1'\|\boldsymbol{\alpha}\|_2\leqslant\|\boldsymbol{\alpha}\|_a\leqslant c_2'\|\boldsymbol{\alpha}\|_2$$

和

$$c_1''\|\boldsymbol{\alpha}\|_b\leqslant\|\boldsymbol{\alpha}\|_2\leqslant c_2''\|\boldsymbol{\alpha}\|_b$$

成立,则显然有

$$c_1'c_1''\|\boldsymbol{\alpha}\|_b\leqslant\|\boldsymbol{\alpha}\|_a\leqslant c_2'c_2''\|\boldsymbol{\alpha}\|_b$$

令 $c_1 = c_1'c_1''$, $c_2 = c_2'c_2''$，便可得定义 5.1.2 中的不等式. 因此，只要对 $b=2$ 证明即可.

考察

$$\|\boldsymbol{\alpha}\|_a = \|x_1\boldsymbol{e}_1 + x_2\boldsymbol{e}_2 + \cdots + x_n\boldsymbol{e}_n\|_a$$

上式可以看成是 $n$ 个坐标分量$(x_1, x_2, \cdots, x_n)$的函数，记

$$\|\boldsymbol{\alpha}\|_a = \varphi(x_1, x_2, \cdots, x_n)$$

可以证明 $\varphi(x_1, x_2, \cdots, x_n)$是坐标分量 $x_1, x_2, \cdots, x_n$ 的连续函数.

事实上，设另一向量为

$$\boldsymbol{\beta} = x_1'\boldsymbol{e}_1 + x_2'\boldsymbol{e}_2 + \cdots + x_n'\boldsymbol{e}_n$$

$$\|\boldsymbol{\beta}\|_a = \varphi(x_1', x_2', \cdots, x_n')$$

则

$$\begin{aligned}
&|\varphi(x_1', x_2', \cdots, x_n') - \varphi(x_1, x_2, \cdots, x_n)| \\
&= |\ \|\boldsymbol{\beta}\|_a - \|\boldsymbol{\alpha}\|_a\ | \leqslant \|\boldsymbol{\beta} - \boldsymbol{\alpha}\|_a \\
&= \|(x_1' - x_1)\boldsymbol{e}_1 + (x_2' - x_2)\boldsymbol{e}_2 + \cdots + (x_n' - x_n)\boldsymbol{e}_n\|_a \\
&\leqslant |x_1' - x_1|\ \|\boldsymbol{e}_1\|_a + |x_2' - x_2|\ \|\boldsymbol{e}_2\|_a + \cdots + |x_n' - x_n|\ \|\boldsymbol{e}_n\|_a
\end{aligned}$$

由于 $\|\boldsymbol{e}_i\|_a (i = 1, 2, \cdots, n)$ 是常数，因此当 $x_i'$ 与 $x_i$ 充分接近时，$\varphi(x_1', x_2', \cdots, x_n')$就充分接近 $\varphi(x_1, x_2, \cdots, x_n)$，这就证明了 $\varphi(x_1, x_2, \cdots, x_n)$是连续函数.

根据连续函数的性质可知，在有界闭集

$$x_1^2 + x_2^2 + \cdots + x_n^2 = 1$$

上(即 $n$ 维欧氏空间 $\mathbf{R}^n$ 的单位球)，函数 $\varphi(x_1, x_2, \cdots, x_n)$可以达到最大值 $M$ 和最小值 $m$，因此 $m > 0$，记

$$d = \|\boldsymbol{\alpha}\|_2 = \left(\sum_{i=1}^{n} x_i^2\right)^{\frac{1}{2}}$$

则向量 $\boldsymbol{\beta} = \frac{x_1}{d}\boldsymbol{e}_1 + \frac{x_2}{d}\boldsymbol{e}_2 + \cdots + \frac{x_n}{d}\boldsymbol{e}_n$ 的分量满足

$$\left(\frac{x_1}{d}\right)^2 + \left(\frac{x_2}{d}\right)^2 + \cdots + \left(\frac{x_n}{d}\right)^2 = 1$$

因此，向量 $\boldsymbol{\beta}$ 在单位球上，从而有

$$0 < m \leqslant \|\boldsymbol{\beta}\|_a = \varphi\left(\frac{x_1}{d}, \frac{x_2}{d}, \cdots, \frac{x_n}{d}\right) \leqslant M$$

但 $\boldsymbol{\beta} = \frac{1}{d}\boldsymbol{\alpha}$，故

$$md \leqslant \|\boldsymbol{\alpha}\|_a \leqslant Md$$

即

$$m\|\boldsymbol{\alpha}\|_2 \leqslant \|\boldsymbol{\alpha}\|_a \leqslant M\|\boldsymbol{\alpha}\|_2$$

即 $\|\boldsymbol{\alpha}\|_a$ 与 $\|\boldsymbol{\alpha}\|_b$ 等价. □

由例 5.1.2～例 5.1.4 易知下列不等式成立

$$\|\boldsymbol{\alpha}\|_\infty \leqslant \|\boldsymbol{\alpha}\|_1 \leqslant n\|\boldsymbol{\alpha}\|_\infty$$

$$\|\boldsymbol{\alpha}\|_\infty \leqslant \|\boldsymbol{\alpha}\|_2 \leqslant \sqrt{n}\|\boldsymbol{\alpha}\|_\infty$$

可见 $\|\boldsymbol{\alpha}\|_1$，$\|\boldsymbol{\alpha}\|_2$，$\|\boldsymbol{\alpha}\|_\infty$互相等价.

注意，定理 5.1.2 不能推广到无穷维线性空间. 另外，若在一种范数意义下向量序列按坐标收敛，则在任何一种范数意义下该向量序列收敛，并且具有相同的极限.

**定理 5.1.3** 在 $\mathbf{C}^n$中，$\lim\limits_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}^* \Leftrightarrow \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\| \to 0$ (当 $k\to\infty$时). 其中 $\|\cdot\|$ 为向量的任意一种范数.

**证** 利用范数的等价性，易知只要对一种范数定理成立，则对任何一种范数都成立，为此取 $\|\cdot\| = \|\cdot\|_\infty$. 显然

$$\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}^* \Leftrightarrow \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_\infty \to 0, \quad k\to\infty$$

而对于 $\mathbf{C}^n$ 上任何一种范数 $\|\cdot\|$，由定理 5.1.2 知，存在常数 $C_1$、$C_2 > 0$，使

$$C_1\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_\infty \leqslant \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\| \leqslant C_2\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_\infty$$

于是有

$$\|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\|_\infty \to 0 \Leftrightarrow \|\boldsymbol{x}^{(k)} - \boldsymbol{x}^*\| \to \infty$$

当 $k\to\infty$时，结论得证. □

如上所述，正因为有了各种范数的等价性，才保证了在各种范数下考虑向量序列收敛的一致性. 因此，我们常常根据不同的要求选择一种方便的范数来研究收敛性问题.

## 5.2 矩阵范数

本节进一步将范数的概念推广到矩阵空间上. 矩阵空间 $\mathbf{C}^{m\times n}$是一个 $mn$ 维线性空间,一个 $m\times n$ 矩阵可以看做一个 $mn$ 维向量,所以可以按向量范数的办法来定义矩阵范数. 但是,矩阵还有矩阵的乘法运算,它应该在定义矩阵范数时予以体现,因此必须多一条反映矩阵乘法的公理. 本节只涉及方阵的范数,至于不是方阵的范数可类似的定义.

**定义 5.2.1** 若对任意的 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$, 都有一个实数 $\|\boldsymbol{A}\|$ 与之对应,且满足下列 4 个条件:

(i) 正定性. 若 $\boldsymbol{A}\neq\mathbf{0}$, 则

$$\|\boldsymbol{A}\|>0$$

(ii) 齐次性. 对任意的 $k\in\mathbf{C}$, 有

$$\|k\boldsymbol{A}\|=|k|\ \|\boldsymbol{A}\|$$

(iii) 三角不等式. 对任意的 $\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{n\times n}$, 有

$$\|\boldsymbol{A}+\boldsymbol{B}\|\leqslant\|\boldsymbol{A}\|+\|\boldsymbol{B}\|$$

(iv) 相容性. 对任意的 $\boldsymbol{A},\boldsymbol{B}\in\mathbf{C}^{n\times n}$, 有

$$\|\boldsymbol{AB}\|\leqslant\|\boldsymbol{A}\|\ \|\boldsymbol{B}\|$$

则称 $\|\boldsymbol{A}\|$ 为 $\mathbf{C}^{n\times n}$上矩阵 $\boldsymbol{A}$ 的矩阵范数.

**例 5.2.1** 设 $\boldsymbol{A}=(a_{ij})\in\mathbf{C}^{n\times n}$, 定义

$$\|\boldsymbol{A}\|_{m_1}=\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|$$

则 $\|\boldsymbol{A}\|_{m_1}$是 $\mathbf{C}^{n\times n}$上的一种矩阵范数,称为矩阵的 $m_1$ -范数.

**证** (i) 正定性、(ii) 齐次性显然.

(iii) 三角不等式. 设 $\boldsymbol{B}=(b_{ij})_{n\times n}$, 则

$$\begin{aligned}\|\boldsymbol{A}+\boldsymbol{B}\|_{m_1}&=\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}+b_{ij}|\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}(|a_{ij}|+|b_{ij}|)\\&\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|+\sum_{i=1}^{n}\sum_{j=1}^{n}|b_{ij}|\\&=\|\boldsymbol{A}\|_{m_1}+\|\boldsymbol{B}\|_{m_1}\end{aligned}$$

(iv) 相容性.
$$AB=\Big(\sum_{k=1}^{n}a_{ik}b_{kj}\Big)_{n\times n}$$

$$\begin{aligned}\|AB\|_{m_1}&=\sum_{i=1}^{n}\sum_{j=1}^{n}\Big|\sum_{k=1}^{n}a_{ik}b_{kj}\Big|\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}\Big(\sum_{k=1}^{n}|a_{ik}||b_{kj}|\Big)\\&\leqslant\sum_{i=1}^{n}\sum_{j=1}^{n}\Big[\Big(\sum_{k=1}^{n}|a_{ik}|\Big)\Big(\sum_{k=1}^{n}|b_{kj}|\Big)\Big]\\&=\sum_{i=1}^{n}\Big(\sum_{k=1}^{n}|a_{ik}|\Big)\sum_{j=1}^{n}\Big(\sum_{k=1}^{n}|b_{kj}|\Big)\\&=\Big(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|\Big)\Big(\sum_{j=1}^{n}\sum_{k=1}^{n}|b_{kj}|\Big)\\&=\|A\|_{m_1}\|B\|_{m_1}\end{aligned}$$

故 $\|A\|_{m_1}$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数.

**例 5.2.2** 设 $A=(a_{ij})\in\mathbf{C}^{n\times n}$，定义

$$\begin{aligned}\|A\|_F=\|A\|_{m_2}&=\sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2}\\&=\Big(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|^2\Big)^{\frac{1}{2}}=\sqrt{\mathrm{tr}(A^{\mathrm{H}}A)}\end{aligned}$$

则 $\|A\|_F$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数，称为矩阵的 Frobenius(弗罗贝尼乌斯)范数，简称 F-范数.

**证** 正定性、齐次性、三角不等式易证. 下证相容性：

设 $B=(b_{ij})\in\mathbf{C}^{n\times n}$，则

$$\begin{aligned}\|AB\|_F&=\Big(\sum_{i=1}^{n}\sum_{j=1}^{n}\Big|\sum_{k=1}^{n}a_{ik}b_{kj}\Big|^2\Big)^{\frac{1}{2}}\\&\leqslant\Big[\sum_{i=1}^{n}\sum_{j=1}^{n}\Big(\sum_{k=1}^{n}|a_{ik}||b_{kj}|^2\Big)\Big]^{\frac{1}{2}}\\&\leqslant\Big\{\sum_{i=1}^{n}\sum_{j=1}^{n}\Big[\Big(\sum_{k=1}^{n}|a_{ik}|^2\Big)\Big(\sum_{k=1}^{n}|b_{kj}|^2\Big)\Big]\Big\}^{\frac{1}{2}}\\&=\Big(\sum_{i=1}^{n}\sum_{k=1}^{n}|a_{ik}|^2\Big)^{\frac{1}{2}}\Big(\sum_{j=1}^{n}\sum_{k=1}^{n}|b_{kj}|^2\Big)^{\frac{1}{2}}\\&=\|A\|_F\|B\|_F\end{aligned}$$

故 $\|A\|_F$ 是 $\mathbf{C}^{n\times n}$ 上的一种范数.

**例 5.2.3** 设 $\boldsymbol{A}=(a_{ij})\in\mathbf{C}^{n\times n}$，定义

$$\|\boldsymbol{A}\|_{m_\infty}=n\max_{i,j}|a_{ij}|$$

则 $\|\boldsymbol{A}\|_{m_\infty}$ 是 $\mathbf{C}^{n\times n}$ 上的矩阵范数，称为矩阵的 $m_\infty$-范数.

**证** (i) 正定性、(ii) 齐次性显然.

(iii) 三角不等式.

设

$$\boldsymbol{B}=(b_{ij})\in\mathbf{C}^{n\times n}$$

则

$$\begin{aligned}\|\boldsymbol{A}+\boldsymbol{B}\|_{m_\infty}&=n\max_{i,j}|a_{ij}+b_{ij}|\leqslant n\max_{i,j}(|a_{ij}|+|b_{ij}|)\\&\leqslant n(\max_{i,j}|a_{ij}+\max_{i,j}|b_{ij}|)\\&=n\max_{i,j}|a_{ij}|+n\max_{i,j}|b_{ij}|\\&=\|\boldsymbol{A}\|_{m_\infty}+\|\boldsymbol{B}\|_{m_\infty}\end{aligned}$$

(iv) 相容性.

$$\begin{aligned}\|\boldsymbol{AB}\|_{m_\infty}&=n\max_{i,j}\left|\sum_{k=1}^{n}a_{ik}b_{kj}\right|\leqslant n\max_{i,j}\sum_{k=1}^{n}|a_{ik}||b_{kj}|\\&\leqslant n\max_{i,j}\sum_{k=1}^{n}(\max_{i,j}|a_{ij}|)|b_{kj}|\\&=n(\max_{i,j}|a_{ij}|)\left(\max_{i,j}\sum_{i=1}^{n}|b_{ij}|\right)\\&\leqslant n\max_{i,j}|a_{ij}|\cdot n\max_{i,j}|b_{ij}|\\&=\|\boldsymbol{A}\|_{m_\infty}\|\boldsymbol{B}\|_{m_\infty}\end{aligned}$$

故 $\|\boldsymbol{A}\|_{m_\infty}$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数.

**例 5.2.4** 设矩阵

$$\boldsymbol{A}=\begin{pmatrix}-2&1&0\\1&-2&0\\3&0&-1\end{pmatrix}$$

试求 $\|\boldsymbol{A}\|_{m_1}$，$\|\boldsymbol{A}\|_F$，$\|\boldsymbol{A}\|_{m_\infty}$.

**解** $$\|\boldsymbol{A}\|_{m_1}=\sum_{i=1}^{3}\sum_{j=1}^{3}|a_{ij}|=2+1+1+2+3+1=10$$

$$\| A \|_F = \left( \sum_{i=1}^{3} \sum_{j=1}^{3} | a_{ij} |^2 \right)^{\frac{1}{2}} = \sqrt{20} = 2\sqrt{5}$$

$$\| A \|_{m_\infty} = 3 \cdot \max_{i,j} | a_{ij} | = 3 \times 3 = 9$$

**例 5.2.5**　设 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$, $n > 1$. 判断

$$\| A \| = \max_{i,j} | a_{ij} |$$

是否为 $\mathbf{C}^{n \times n}$ 上的矩阵范数.

**解**　取

$$A_0 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}, \qquad B_0 = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 1 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \cdots & 0 \end{pmatrix}$$

则 $\| A_0 \| = 1$, $\| B_0 \| = 1$, 但

$$A_0 B_0 = \begin{pmatrix} n & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

从而 $\| A_0 B_0 \| = n$. 由于 $n > 1$,所以 $\| A_0 B_0 \| > \| A_0 \| \ \| B_0 \|$,不满足矩阵范数中的相容性条件,故不是矩阵范数.

与向量的情形一样,矩阵也有各种各样的范数,而且在大多数情况下,矩阵范数和向量范数混合在一起使用. 因此我们考虑的一些矩阵范数,应当将它们与向量范数联系起来.

**定义 5.2.2**　设 $A \in \mathbf{C}^{n \times n}$, $\alpha \in \mathbf{C}^n$, 如果取定的向量范数 $\| \alpha \|$ 和矩阵范数 $\| A \|$ 满足不等式

$$\| A\alpha \| \leqslant \| A \| \| \alpha \|$$

则称矩阵范数 $\| A \|$ 与向量范数 $\| \alpha \|$ 是相容的.

**例 5.2.6**　证明 $\mathbf{C}^{n \times n}$ 上的矩阵 $m_1$-范数与 $\mathbf{C}^n$ 上的向量 1-范数相容.

**证**　设 $A = (a_{ij}) \in \mathbf{C}^{n \times n}$, $\alpha = (x_1, x_2, \cdots, x_n)^T \in \mathbf{C}^n$, 则

$$\| A\alpha \|_1 = \sum_{i=1}^{n} \left| \sum_{j=1}^{n} a_{ij} x_j \right| \leqslant \sum_{i=1}^{n} \left( \sum_{j=1}^{n} | a_{ij} | | x_j | \right)$$

$$\leqslant \sum_{i=1}^{n} \left[ \left( \sum_{j=1}^{n} | a_{ij} | \right) \left( \sum_{j=1}^{n} | x_j | \right) \right]$$

$$=\left(\sum_{i=1}^{n}\sum_{j=1}^{n}|a_{ij}|\right)\sum_{j=1}^{n}|x_j|=\|\mathbf{A}\|_{m_1}\|\boldsymbol{\alpha}\|_1$$

即证明了矩阵的 $m_1$ -范数与向量的 1 -范数相容.

类似地还可以证明：$\mathbf{C}^{n\times n}$ 上矩阵的 $m_\infty$-范数与 $\mathbf{C}^n$ 上向量的 1 -范数、2 -范数均相容. 还可以证明：

(1) $\mathbf{C}^{n\times n}$ 上的每一种矩阵范数，在 $\mathbf{C}^n$ 上都存在与该矩阵范数相容的向量范数.

(2) $\mathbf{C}^{n\times n}$ 上任意两种矩阵范数 $\|\mathbf{A}\|_\alpha$，$\|\mathbf{A}\|_\beta$ 都是等价的，即存在两个与 $\mathbf{A}$ 无关的正数 $C_1$、$C_2$，使得

$$C_1\|\mathbf{A}\|_\beta\leqslant\|\mathbf{A}\|_\alpha\leqslant C_2\|\mathbf{A}\|_\beta$$

除上述介绍的几种矩阵范数外，下列三种也是比较常用的矩阵范数：

(1) $\|\mathbf{A}\|_1=\max\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^{n}|a_{ij}|$（列模和最大者，也称为列范数）.

(2) $\|\mathbf{A}\|_\infty=\max\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^{n}|a_{ij}|$（行模和最大者，也称为行范数）.

(3) $\|\mathbf{A}\|_2=\sqrt{\lambda_n}=\sqrt{\max\limits_i\lambda_i(\mathbf{A}^{\mathrm{H}}\mathbf{A})}$（$\lambda_n$ 是矩阵 $\mathbf{A}^{\mathrm{H}}\mathbf{A}$ 的最大特征值，也称为谱范数）.

## 5.3 范数的应用

在前两节中. 我们已对向量和矩阵的范数作了初步介绍. 本节介绍矩阵范数的一些应用，以便为读者进一步的学习和应用打下基础.

**定义 5.3.1** 设 $\mathbf{A}\in\mathbf{C}^{n\times n}$，$\lambda_1, \lambda_2, \cdots, \lambda_j, \cdots, \lambda_n$ 是 $\mathbf{A}$ 的 $n$ 个特征值，称

$$\rho(\mathbf{A})=\max_j|\lambda_j|$$

为 $\mathbf{A}$ 的谱半径.

关于矩阵的谱半径有如下定理.

**定理 5.3.1** 设 $\mathbf{A}\in\mathbf{C}^{n\times n}$，则

(i) $\rho(\mathbf{A}^k)=[\rho(\mathbf{A})]^k$；

(ii) $\rho(\mathbf{A}^{\mathrm{H}}\mathbf{A})=\rho(\mathbf{A}\mathbf{A}^{\mathrm{H}})=(\|\mathbf{A}\|_2)^2$.

**证** (i) 设 $\mathbf{A}$ 的 $n$ 个特征值为 $\lambda_1, \lambda_2, \cdots, \lambda_j, \cdots, \lambda_n$，则 $\mathbf{A}^k$ 的特征值为 $\lambda_1^k, \lambda_2^k, \cdots, \lambda_j^k, \cdots, \lambda_n^k$，于是

$$\rho(\mathbf{A}^k)=\max_j|\lambda_j^k|=(\max_j|\lambda_j|)^k=[\rho(\mathbf{A})]^k$$

(ii) 由矩阵 2 -范数计算公式知

$$\| A \|_2 = \sqrt{\lambda_n} = \sqrt{\max_j \lambda_j (A^H A)}$$

又因 $A^H A$ 与 $AA^H$ 有相同的非零特征值,故

$$\rho(A^H A) = \rho(AA^H) = (\| A \|_2)^2$$

□

有了谱半径的概念,可以对矩阵范数作如下的初步估计.

**定理 5.3.2**　设 $A \in C^{n\times n}$, 则对 $C^{n\times n}$ 上的任意一种范数 $\| \cdot \|$, 都有

$$\rho(A) \leqslant \| A \|$$

**证**　设 $\lambda$ 为 $A$ 的特征值, $\alpha$ 为 $A$ 的属于特征值 $\lambda$ 的特征向量, 故 $\alpha \neq 0$, 即 $\| \alpha \| \neq 0$. 另设 $\| \cdot \|_v$ 是 $C^n$ 上与矩阵范数 $\| \cdot \|$ 相容的向量范数, 由 $A\alpha = \lambda\alpha$ 应有

$$\| A\alpha \|_v = | \lambda | \; \| \alpha \|_v$$

而

$$\| A\alpha \|_v \leqslant \| A \| \; \| \alpha \|_v$$

于是有

$$| \lambda | \; \| \alpha \|_v \leqslant \| A \| \; \| \alpha \|_v$$

同除 $\| \alpha \|_v (\neq 0)$, 有

$$| \lambda | \leqslant \| A \|$$

于是

$$\rho(A) \leqslant \| A \|$$

□

定理 5.3.2 为我们估计矩阵范数提供了一个下限. 利用矩阵的 Jordan 标准形,在某种程度上可以认为给出矩阵范数的一个上限.

**定理 5.3.3**　设 $A \in C^{n\times n}$, $\forall \varepsilon > 0$, 则存在某一矩阵范数 $\| \cdot \|_m$, 使得

$$\| A \|_m \leqslant \rho(A) + \varepsilon$$

**证**　由定理 3.1.1 知, $\forall A \in C^{n\times n}$, 都存在可逆阵 $P \in C^{n\times n}$, 使

$$P^{-1}AP = J = \begin{pmatrix} \lambda_1 & k_1 & & & \\ & \lambda_2 & k_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & k_{n-1} \\ & & & & \lambda_n \end{pmatrix}$$

其中, $k_i = 0$ 或 1; $i = 1, 2, \cdots, n-1$. 取

$$D = \mathrm{diag}(1, \varepsilon, \varepsilon^2, \cdots, \varepsilon^{n-1})$$

则
$$D^{-1} = \mathrm{diag}\left(1, \frac{1}{\varepsilon}, \frac{1}{\varepsilon^2}, \cdots, \frac{1}{\varepsilon^{n-1}}\right)$$

于是

$$D^{-1}(P^{-1}AP)D = D^{-1}JD = \begin{pmatrix} \lambda_1 & \varepsilon k_1 & & & \\ & \lambda_2 & \varepsilon k_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & \varepsilon k_n \\ & & & & \lambda_n \end{pmatrix}$$

考虑矩阵的∞-范数有

$$\begin{aligned} \| D^{-1}JD \|_\infty &= \max_j \{ |\lambda_j| + \varepsilon k_j \} \quad (\text{约定 } k_n = 0) \\ &\leqslant \max_j \{ |\lambda_j| + \varepsilon \} \\ &= \rho(A) + \varepsilon \end{aligned}$$

对任意的 $A \in \mathbf{C}^{n\times n}$，定义

$$\| A \|_m = \| D^{-1}P^{-1}APD \|_\infty = \| D^{-1}JD \|_\infty$$

可以验证 $\| A \|_m$是 $\mathbf{C}^{n\times n}$上的一种矩阵范数，且

$$\| A \|_m \leqslant \rho(A) + \varepsilon$$ □

在工程实际中，我们往往需要计算 $A^{-1}$ 和求解线性方程组 $Ax = b$，这里 $A \in \mathbf{C}^{n\times n}$，$b \in \mathbf{C}^n$. 现在的问题是$A$，$b$ 的微小变化 $\Delta A$ 和 $\Delta b$ 将会对实际问题的解产生什么影响？对于矩阵求逆而言，研究 $A^{-1}$ 与 $(A+\Delta A)^{-1}$ 是否非常接近；对于线性方程组来说就是解决误差扰动 $\Delta A$，$\Delta b$ 引起解的误差 $\Delta x$ 的大小如何变化等问题.

**定理 5.3.4** 设 $B \in \mathbf{C}^{n\times n}$，若对 $\mathbf{C}^{n\times n}$ 上某一矩阵范数 $\| \cdot \|$，$\| B \| < 1$，则 $E - B$ 可逆.

**证** 反证法.

设 $E - B$ 不可逆，则齐次线性方程组$(E - B)x = 0$ 有非零解 $\alpha$，使

$$(E - B)\alpha = 0$$

即
$$\alpha = B\alpha$$

设 $\| \cdot \|_a$ 是 $\mathbf{C}^n$上与矩阵范数 $\| \cdot \|$ 相容的向量范数，则

$$\| \alpha \|_a = \| B\alpha \|_a \leqslant \| B \| \ \| \alpha \|_a$$

因为 $\|\boldsymbol{\alpha}\|_a > 0$，上式两端除以 $\|\boldsymbol{\alpha}\|_a$ 得 $\|\boldsymbol{B}\| \geqslant 1$，矛盾. 所以 $\boldsymbol{E}-\boldsymbol{B}$ 可逆. □

**定理 5.3.5**　设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$ 且 $\boldsymbol{A}$ 可逆，若对 $\mathbf{C}^{n\times n}$ 上某一矩阵范数 $\|\cdot\|$，如果 $\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\| < 1$，则

(i) $\boldsymbol{A}+\Delta\boldsymbol{A}$ 可逆；

(ii) $\|\boldsymbol{A}+\Delta\boldsymbol{A}\| \leqslant \dfrac{\|\boldsymbol{A}^{-1}\|}{1-\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\|}$；

(iii) $\dfrac{\|\boldsymbol{A}^{-1}-(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}\|}{\|\boldsymbol{A}^{-1}\|} \leqslant \dfrac{\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\|}{1-\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\|}$.

**证**　(i)

$$\begin{aligned}\boldsymbol{A}+\Delta\boldsymbol{A} &= \boldsymbol{A}\boldsymbol{E}+\boldsymbol{A}\boldsymbol{A}^{-1}\Delta\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{E}+\boldsymbol{A}^{-1}\Delta\boldsymbol{A})\\ &= \boldsymbol{A}[\boldsymbol{E}-(-\boldsymbol{A}^{-1}\Delta\boldsymbol{A})]\end{aligned}$$

因为

$$\|-\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\| = \|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\| < 1$$

由定理 5.3.4 知 $\boldsymbol{E}-(-\boldsymbol{A}^{-1}\Delta\boldsymbol{A})$ 可逆，故 $\boldsymbol{A}+\Delta\boldsymbol{A}$ 可逆.

(ii) 因 $\boldsymbol{A}+\Delta\boldsymbol{A}$ 可逆，故

$$(\boldsymbol{A}+\Delta\boldsymbol{A})(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1} = \boldsymbol{E}$$

将前一个括号乘开有

$$\boldsymbol{A}(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}+\Delta\boldsymbol{A}(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1} = \boldsymbol{E}$$

$$\boldsymbol{A}(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1} = \boldsymbol{E}-\Delta\boldsymbol{A}(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}$$

左乘 $\boldsymbol{A}^{-1}$，有

$$(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1} = \boldsymbol{A}^{-1}-\boldsymbol{A}^{-1}\Delta\boldsymbol{A}(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}$$

由范数的三角不等式得

$$\|(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}\| \leqslant \|\boldsymbol{A}^{-1}\| + \|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\|\,\|(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}\|$$

右端第二项移项得

$$\|(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}\|(1-\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\|) \leqslant \|\boldsymbol{A}^{-1}\|$$

由已知 $1-\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\| > 0$，除以 $1-\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\|$ 得

$$\|(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}\| \leqslant \frac{\|\boldsymbol{A}^{-1}\|}{1-\|\boldsymbol{A}^{-1}\Delta\boldsymbol{A}\|}$$

(iii)

$$\begin{aligned}\boldsymbol{A}^{-1}-(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1} &= \boldsymbol{A}^{-1}(\boldsymbol{A}+\Delta\boldsymbol{A})(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}-\boldsymbol{A}^{-1}\boldsymbol{A}(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}\\ &= \boldsymbol{A}^{-1}[(\boldsymbol{A}+\Delta\boldsymbol{A})-\boldsymbol{A}](\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}\\ &= \boldsymbol{A}^{-1}\Delta\boldsymbol{A}(\boldsymbol{A}+\Delta\boldsymbol{A})^{-1}\end{aligned}$$

两端取范数，根据相容性及(ii)得

$$\begin{aligned}\|A^{-1}-(A+\Delta A)^{-1}\| &\leqslant \|A^{-1}\Delta A\|\ \|(A+\Delta A)^{-1}\| \\ &\leqslant \|A^{-1}\Delta A\| \frac{\|A^{-1}\|}{1-\|A^{-1}\Delta A\|}\end{aligned}$$

此即

$$\frac{\|A^{-1}-(A+\Delta A)^{-1}\|}{\|A^{-1}\|}\leqslant\frac{\|A^{-1}\|}{1-\|A^{-1}\Delta A\|}$$

□

**推论 5.3.1** 设 $A\in C^{n\times n}$ 且 $A$ 可逆，$\Delta A\in C^{n\times n}$，对 $C^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$，若 $\|A^{-1}\|\ \|\Delta A\|<1$，则

$$\frac{\|A^{-1}-(A+\Delta A)^{-1}\|}{\|A^{-1}\|}\leqslant\frac{\|A\|\ \|A^{-1}\|\ \dfrac{\|\Delta A\|}{\|A\|}}{1-\|A\|\ \|A^{-1}\|\ \dfrac{\|\Delta A\|}{\|A\|}}$$

关于线性方程组 $Ax=b$，矩阵及解的扰动有下述定理.

**定理 5.3.6** 设 $A\in C^{n\times n}$ 且 $A$ 可逆，$\Delta A\in C^{n\times n}$，$b$，$\Delta b\in C^{n}$，对 $C^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$，若 $\|A^{-1}\|\ \|\Delta A\|<1$，则线性方程组

$$Ax=b \text{ 与 } (A+\Delta A)(x+\Delta x)=b+\Delta b$$

的解满足

$$\frac{\|\Delta x\|_a}{\|x\|_a}\leqslant\frac{\|A\|\ \|A^{-1}\|}{1-\|A\|\ \|A^{-1}\|\ \dfrac{\|\Delta A\|}{\|A\|}}\left(\frac{\|\Delta A\|}{\|A\|}+\frac{\|\Delta b\|_a}{\|b\|_a}\right)$$

其中 $\|\cdot\|_a$ 是 $C^n$ 上与矩阵范数 $\|\cdot\|$ 相容的向量范数.

**证**

$$(A+\Delta A)(x+\Delta x)=b+\Delta b$$

$$Ax+(\Delta A)x+A(\Delta x)+(\Delta A)(\Delta x)=b+\Delta b$$

由于 $Ax=b$，则

$$A(\Delta x)+(\Delta A)x+(\Delta A)(\Delta x)=\Delta b$$

左乘 $A^{-1}$ 得

$$\Delta x+A^{-1}(\Delta A)x+A^{-1}(\Delta A)(\Delta x)=A^{-1}(\Delta b)$$

即

$$\Delta x = -A^{-1}(\Delta A)x - A^{-1}(\Delta A)(\Delta x) + A^{-1}(\Delta b)$$

取范数，由三角不等式及相容性有

$$\|\Delta x\|_a \leqslant \|A^{-1}\| \|\Delta A\| \|x\|_a + \|A^{-1}\| \|\Delta A\| \|\Delta x\|_a + \|A^{-1}\| \|\Delta b\|_a$$

移项得

$$\|\Delta x\|_a(1 - \|A^{-1}\| \|\Delta x\|) \leqslant \|A^{-1}\|(\|\Delta A\| \|x\|_a + \|\Delta b\|_a)$$

由已知条件知 $1 - \|A^{-1}\| \|\Delta A\| > 0$，上式除以 $\|x\|_a(1 - \|A^{-1}\| \|\Delta A\|)$ 得

$$\frac{\|\Delta x\|_a}{\|x\|_a} \leqslant \frac{\|A^{-1}\|}{1 - \|A^{-1}\| \|\Delta A\|}\left(\|\Delta A\| + \frac{\|\Delta b\|_a}{\|x\|_a}\right)$$

$$= \frac{\|A\| \|A^{-1}\|}{1 - \|A^{-1}\| \|\Delta A\|}\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|_a}{\|A\| \|x\|_a}\right)$$

以 $\|b\|_a = \|Ax\|_a \leqslant \|A\| \|x\|_a$ 代入上式右端括号中第二项的分母得

$$\frac{\|\Delta x\|_a}{\|x\|_a} \leqslant \frac{\|A\| \|A^{-1}\|}{1 - \|A\| \|A^{-1}\| \dfrac{\|\Delta A\|}{\|A\|}}\left(\frac{\|\Delta A\|}{\|A\|} + \frac{\|\Delta b\|_a}{\|b\|_a}\right) \qquad \square$$

从上述式子可以看出，在计算逆矩阵和线性方程组的求解中，误差扰动 $\Delta A$，$\Delta b$，$\Delta x$ 对其影响与 $\|A\| \|A^{-1}\|$ 的大小有密切关系.

**定义 5.3.2**　设 $A \in \mathbf{C}^{n\times n}$，且 $A$ 为非奇异矩阵，称数

$$\mathrm{Cond}(A) = \|A\| \|A^{-1}\|$$

为矩阵 $A$ 的条件数.

一般地，若 Cond($A$)较大，就说 $A$ 对于求逆或求解线性方程组是病态的或不稳定的，否则称为良态的或稳定的. 究竟条件数多大矩阵才算病态，一般来讲是没有具体标准的，也只是相对而言.

通常使用的条件数有

(i) $\mathrm{Cond}(A)_\infty = \|A\|_\infty \|A^{-1}\|_\infty$；

(ii) $A$ 的谱条件数

$$\mathrm{Cond}(A)_2 = \|A\|_2 \|A^{-1}\|_2 = \sqrt{\frac{\lambda_{\max}(A^H A)}{\lambda_{\min}(A^H A)}}$$

**例 5.3.1**　设 $A = \begin{pmatrix} 1 & 10^4 \\ 1 & 1 \end{pmatrix}$，计算 $\mathrm{Cond}(A)_\infty$.

**解**

$$A^{-1} = \frac{1}{10^4 - 1}\begin{pmatrix} -1 & 10^4 \\ 1 & -1 \end{pmatrix}$$

$$\text{Cond}(A)_\infty = \frac{(1+10^4)^2}{10^4 - 1} \approx 10^4$$

当矩阵 $A$ 的元素大小不均匀时，往往造成很大的条件数，在这种情况下，如果要解方程组，则可以对 $A$ 的行引进适当的比例因子，以减小条件数.

**例 5.3.2** 设有方程组

$$\begin{pmatrix} 1 & 10^4 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 10^4 \\ 2 \end{pmatrix}$$

在 $A$ 的第一行引进比例因子，若用 $r_1 = \max\limits_{1\leqslant j\leqslant 2} |a_{1j}| = 10^4$ 除第一个方程，则得 $A'x = b'$，即

$$\begin{pmatrix} 10^{-4} & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$(A')^{-1} = \frac{1}{1-10^{-4}}\begin{pmatrix} -1 & 1 \\ 1 & -10^4 \end{pmatrix}$$

于是

$$\text{Cond}(A')_\infty = \frac{4}{1-10^{-4}} \approx 4$$

下面我们介绍范数的另一个应用. 由第 2 章的讨论已知，对不相容线性方程组 $Ax = b$ 可以求其最小二乘解. 一般地说，该线性方程组的最小二乘解不是唯一的. 在众多的最小二乘解中 2-范数为最小者.

**定义 5.3.3** 假定 $A \in \mathbf{C}^{m\times n}$，$Ax = b$ 是不相容线性方程组，若有向量 $x_0 \in \mathbf{C}^n$ 使得

$$\| Ax_0 - b \| = \min_{x\in \mathbf{C}^n} \| Ax - b \|$$

则称 $x_0$ 为不相容线性方程组 $Ax = b$ 的一个近似解，或最小二乘解，记为 LNLS 解.

在 1.1 节中，我们已经知道

$$\text{Im}(A) = \{Ax \mid x \in \mathbf{C}^n\} \subset \mathbf{C}^m$$

是一个子空间，根据定理 2.4.2，若设 $b_0$ 是 $b$ 在 $\text{Im}(A)$ 上的正交投影，则必有

$$\| \boldsymbol{b}_0 - \boldsymbol{b} \| = \min_{x \in \mathbf{C}^n} \| \boldsymbol{A}\boldsymbol{x} - \boldsymbol{b} \| \tag{5.3.1}$$

且满足式(5.3.1)的 $\boldsymbol{b}_0$ 是唯一存在的.

设 $\operatorname{rank}(\boldsymbol{A}) = r$,则有 $\dim \operatorname{Im}(\boldsymbol{A}) = r$. 利用4.4节中的结果,若设 $\lambda_1, \lambda_2, \cdots, \lambda_r$ 是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 的非零特征值, $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r$ 是分别属于这些特征值的标准正交特征向量,则 $s_i = \sqrt{\lambda_i}$, $i = 1, 2, \cdots, r$ 就是 $\boldsymbol{A}$ 的非零奇异值,且

$$\boldsymbol{y}_i = \frac{1}{s_i}\boldsymbol{A}\boldsymbol{x}_i, \quad i = 1, 2, \cdots, r$$

是标准正交向量组. 从而 $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_r\}$ 就是 $\operatorname{Im}(\boldsymbol{A})$ 的标准正交基,于是由定理2.2.2得

$$\begin{aligned} \boldsymbol{b}_0 &= (\boldsymbol{b}, \boldsymbol{y}_1)\boldsymbol{y}_1 + (\boldsymbol{b}, \boldsymbol{y}_2)\boldsymbol{y}_2 + \cdots + (\boldsymbol{b}, \boldsymbol{y}_r)\boldsymbol{y}_r \\ &= (\boldsymbol{y}_1^{\mathrm{H}}\boldsymbol{b})\boldsymbol{y}_1 + (\boldsymbol{y}_2^{\mathrm{H}}\boldsymbol{b})\boldsymbol{y}_2 + \cdots + (\boldsymbol{y}_r^{\mathrm{H}}\boldsymbol{b})\boldsymbol{y}_r \end{aligned}$$

**定理 5.3.7** 设 $\boldsymbol{A} \in \mathbf{C}^{m \times n}$, $\boldsymbol{b} \in \mathbf{C}^m$,则不相容线性方程组 $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ 的全部最小二乘解为

$$\boldsymbol{x} = \boldsymbol{A}^+ \boldsymbol{b} + (\boldsymbol{E} - \boldsymbol{A}^+ \boldsymbol{A})\boldsymbol{C} \tag{5.3.2}$$

其中 $\boldsymbol{C} \in \mathbf{C}^n$ 是任意列向量.

**证** 证明分两部分.

首先证明,若取 $\boldsymbol{x}_0 = \boldsymbol{A}^+ \boldsymbol{b}$,则有 $\boldsymbol{A}\boldsymbol{x}_0 = \boldsymbol{b}_0$,从而 $\boldsymbol{A}^+\boldsymbol{b}$ 是一个最小二乘解. 事实上,根据4.5节中定理4.5.3的证明,取 $\boldsymbol{V}_1 = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r)$, $\boldsymbol{U}_1 = (\boldsymbol{y}_1, \boldsymbol{y}_2, \cdots, \boldsymbol{y}_r)$, $D = \operatorname{diag}(d_1, d_2, \cdots, d_r)$ 时

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x}_0 &= \boldsymbol{A}\boldsymbol{A}^+ \boldsymbol{b} = \boldsymbol{A}(\boldsymbol{V}_1\boldsymbol{D}^{-1}\boldsymbol{U}_1^{\mathrm{H}})\boldsymbol{b} \\ &= \boldsymbol{A}\left[ (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_r) \begin{pmatrix} d_1^{-1} & & & \\ & d_2^{-1} & & \\ & & \ddots & \\ & & & d_r^{-1} \end{pmatrix} \begin{pmatrix} y_1^{\mathrm{H}} \\ y_2^{\mathrm{H}} \\ \vdots \\ y_r^{\mathrm{H}} \end{pmatrix} \boldsymbol{b} \right] \\ &= \boldsymbol{A}[(d_1^{-1}\boldsymbol{y}_1^{\mathrm{H}}\boldsymbol{b})\boldsymbol{x}_1 + (d_2^{-1}\boldsymbol{y}_2^{\mathrm{H}}\boldsymbol{b})\boldsymbol{x}_2 + \cdots + (d_r^{-1}\boldsymbol{y}_r^{\mathrm{H}}\boldsymbol{b})\boldsymbol{x}_r] \\ &= (d_1^{-1}\boldsymbol{y}_1^{\mathrm{H}}\boldsymbol{b})\boldsymbol{A}\boldsymbol{x}_1 + (d_2^{-1}\boldsymbol{y}_2^{\mathrm{H}}\boldsymbol{b})\boldsymbol{A}\boldsymbol{x}_2 + \cdots + (d_r^{-1}\boldsymbol{y}_r^{\mathrm{H}}\boldsymbol{b})\boldsymbol{A}\boldsymbol{x}_r \\ &= (\boldsymbol{y}_1^{\mathrm{H}}\boldsymbol{b})\boldsymbol{y}_1 + (\boldsymbol{y}_2^{\mathrm{H}}\boldsymbol{b})\boldsymbol{y}_2 + \cdots + (\boldsymbol{y}_r^{\mathrm{H}}\boldsymbol{b})\boldsymbol{y}_r \\ &= \boldsymbol{b}_0 \end{aligned}$$

其次证明,任一最小二乘解 $\boldsymbol{x}_1$ 必可以表示成式(5.3.2)的形式.

假定 $\boldsymbol{x}_1$ 是不相容线性方程组的一个最小二乘解，由 $\min\limits_{x\in\mathbf{C}^n}\|\boldsymbol{Ax}-\boldsymbol{b}\|$ 的唯一性可知

$$\|\boldsymbol{Ax}_1-\boldsymbol{b}\| = \min_{x\in\mathbf{C}^n}\|\boldsymbol{Ax}-\boldsymbol{b}\| = \|\boldsymbol{b}_0-\boldsymbol{b}\| = \|\boldsymbol{AA}^+\boldsymbol{b}-\boldsymbol{b}\| \tag{5.3.3}$$

因为

$$\|\boldsymbol{Ax}_1-\boldsymbol{b}\| = \|(\boldsymbol{Ax}_1-\boldsymbol{AA}^+\boldsymbol{b})+(\boldsymbol{AA}^+\boldsymbol{b}-\boldsymbol{b})\| \tag{5.3.4}$$

且

$$\begin{aligned}
&(\boldsymbol{Ax}_1-\boldsymbol{AA}^+\boldsymbol{b})^{\mathrm{H}}(\boldsymbol{AA}^+\boldsymbol{b}-\boldsymbol{b})\\
&=[\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}-\boldsymbol{b}^{\mathrm{H}}(\boldsymbol{AA}^+)^{\mathrm{H}}](\boldsymbol{AA}^+\boldsymbol{b}-\boldsymbol{b})\\
&=\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{AA}^+\boldsymbol{b}-\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}-\boldsymbol{b}^{\mathrm{H}}(\boldsymbol{AA}^+)^{\mathrm{H}}(\boldsymbol{AA}^+\boldsymbol{b})+\boldsymbol{b}^{\mathrm{H}}(\boldsymbol{AA}^+)^{\mathrm{H}}\boldsymbol{b}\\
&=\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}(\boldsymbol{AA}^+)^{\mathrm{H}}\boldsymbol{b}-\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}-\boldsymbol{b}^{\mathrm{H}}(\boldsymbol{AA}^+\boldsymbol{AA}^+)\boldsymbol{b}+\boldsymbol{b}^{\mathrm{H}}(\boldsymbol{AA}^+)\boldsymbol{b}\\
&=\boldsymbol{x}_1^{\mathrm{H}}(\boldsymbol{AA}^+\boldsymbol{A})^{\mathrm{H}}\boldsymbol{b}-\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}-\boldsymbol{b}^{\mathrm{H}}\boldsymbol{AA}^+\boldsymbol{b}+\boldsymbol{b}^{\mathrm{H}}\boldsymbol{AA}^+\boldsymbol{b}\\
&=\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}-\boldsymbol{x}_1^{\mathrm{H}}\boldsymbol{A}^{\mathrm{H}}\boldsymbol{b}=\boldsymbol{0}
\end{aligned}$$

所以，$\boldsymbol{Ax}_1-\boldsymbol{AA}^+\boldsymbol{b}$ 与 $\boldsymbol{AA}^+\boldsymbol{b}-\boldsymbol{b}$ 正交，于是根据例 2.1.4 的结果，由式(5.3.4)得

$$\|\boldsymbol{Ax}_1-\boldsymbol{b}\|^2 = \|\boldsymbol{Ax}_1-\boldsymbol{AA}^+\boldsymbol{b}\|^2+\|\boldsymbol{AA}^+\boldsymbol{b}-\boldsymbol{b}\|^2$$

从而由式(5.3.3)知

$$\|\boldsymbol{Ax}_1-\boldsymbol{AA}^+\boldsymbol{b}\| = 0$$

即

$$\boldsymbol{Ax}_1 = \boldsymbol{AA}^+\boldsymbol{b}$$

这说明 $\boldsymbol{x}_1$ 是相容线性方程组 $\boldsymbol{Ax}=\boldsymbol{AA}^+\boldsymbol{b}$ 的解，于是存在 $\boldsymbol{C}_1\in\mathbf{C}^n$ 使得

$$\boldsymbol{x}_1 = \boldsymbol{A}^+\boldsymbol{b}+(\boldsymbol{E}-\boldsymbol{A}^+\boldsymbol{A})\boldsymbol{C}_1$$ □

设 $\boldsymbol{Ax}=\boldsymbol{b}$ 是不相容线性方程组，若有一个最小二乘解 $\boldsymbol{x}_0$，使得

$$\|\boldsymbol{x}_0\| = \min\left\{\|\boldsymbol{x}\| \,\middle|\, \boldsymbol{x}=\boldsymbol{A}^+\boldsymbol{b}+(\boldsymbol{E}-\boldsymbol{A}^+\boldsymbol{A})\boldsymbol{C},\ \boldsymbol{C}\text{任意}\right\}$$

则称 $\boldsymbol{x}_0$ 是 $\boldsymbol{Ax}=\boldsymbol{b}$ 的极小范数最小二乘解.

**定理 5.3.8** 设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，$\boldsymbol{b}\in\mathbf{C}^m$，则不相容线性方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 有唯一的极小范数最小二乘解 $\boldsymbol{x}_0=\boldsymbol{A}^+\boldsymbol{b}$.

**证** 任取一个最小二乘解 $\boldsymbol{x}_1$,则必存在 $\boldsymbol{C}_1 \in \mathbf{C}^n$,使得

$$\boldsymbol{x}_1 = \boldsymbol{A}^+ \boldsymbol{b} + (\boldsymbol{E} - \boldsymbol{A}^+ \boldsymbol{A})\boldsymbol{C}_1 \tag{5.3.5}$$

因为

$$\begin{aligned}(\boldsymbol{A}^+ \boldsymbol{b})^{\mathrm{H}}(\boldsymbol{I} - \boldsymbol{A}^+ \boldsymbol{A})\boldsymbol{C}_1 &= \boldsymbol{b}^{\mathrm{H}}(\boldsymbol{A}^+)^{\mathrm{H}}\boldsymbol{C}_1 - \boldsymbol{b}^{\mathrm{H}}(\boldsymbol{A}^+)^{\mathrm{H}}\boldsymbol{A}^+ \boldsymbol{A}\boldsymbol{C}_1 \\ &= \boldsymbol{b}^{\mathrm{H}}(\boldsymbol{A}^+)^{\mathrm{H}}\boldsymbol{C}_1 - \boldsymbol{b}^{\mathrm{H}}(\boldsymbol{A}^+)^{\mathrm{H}}(\boldsymbol{A}^+ \boldsymbol{A})^{\mathrm{H}}\boldsymbol{C}_1 \\ &= \boldsymbol{b}^{\mathrm{H}}(\boldsymbol{A}^+)^{\mathrm{H}}\boldsymbol{C}_1 - \boldsymbol{b}^{\mathrm{H}}(\boldsymbol{A}^+ \boldsymbol{A}\boldsymbol{A}^+)^{\mathrm{H}}\boldsymbol{C}_1 \\ &= \boldsymbol{b}^{\mathrm{H}}(\boldsymbol{A}^+)^{\mathrm{H}}\boldsymbol{C}_1 - \boldsymbol{b}^{\mathrm{H}}(\boldsymbol{A}^+)^{\mathrm{H}}\boldsymbol{C}_1 \\ &= \boldsymbol{0}\end{aligned}$$

所以

$$\boldsymbol{A}^+ \boldsymbol{b} \perp (\boldsymbol{I} - \boldsymbol{A}^+ \boldsymbol{A})\boldsymbol{C}_1$$

于是由式(5.3.5)得

$$\|\boldsymbol{x}_1\|^2 = \|\boldsymbol{A}^+ \boldsymbol{b}\|^2 + \|(\boldsymbol{E} - \boldsymbol{A}^+ \boldsymbol{A})\boldsymbol{C}_1\|^2 \geqslant \|\boldsymbol{x}_0\|^2$$

上式表明,$\boldsymbol{x}_0 = \boldsymbol{A}^+ \boldsymbol{b}$ 不仅是极小范数最小二乘解,而且还是唯一的. □

**例 5.3.3** 求不相容线性方程组

$$\begin{pmatrix} 1 & 0 & -1 & 1 \\ 0 & 2 & 2 & 2 \\ -1 & 4 & 5 & 3 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix} = \begin{pmatrix} 4 \\ 1 \\ 2 \end{pmatrix}$$

的极小范数最小二乘解.

**解** 由满秩分解得系数矩阵 $\boldsymbol{A}$ 的广义逆为

$$\boldsymbol{A}^+ = \frac{1}{18}\begin{pmatrix} 5 & 2 & -1 \\ 1 & 1 & 1 \\ -4 & -1 & 2 \\ 6 & 3 & 0 \end{pmatrix}$$

于是方程组的极小范数最小二乘解为

$$\boldsymbol{x} = \boldsymbol{A}^+ \boldsymbol{b} = \frac{1}{18}(20,\ 7,\ -13,\ 27)^{\mathrm{T}}$$

# 习题 5

1. 设 $\boldsymbol{\alpha}=(1+\mathrm{i}, -2, 4\mathrm{i}, 1, 0)^{\mathrm{T}}$，试求：$\|\boldsymbol{\alpha}\|_1$，$\|\boldsymbol{\alpha}\|_2$，$\|\boldsymbol{\alpha}\|_\infty$.

2. 证明：若 $\boldsymbol{\alpha}\in\mathbf{C}^n$，则

(1) $\|\boldsymbol{\alpha}\|_2\leqslant\|\boldsymbol{\alpha}\|_1\leqslant\sqrt{n}\|\boldsymbol{\alpha}\|_2$；

(2) $\|\boldsymbol{\alpha}\|_\infty\leqslant\|\boldsymbol{\alpha}\|_1\leqslant n\|\boldsymbol{\alpha}\|_\infty$；

(3) $\|\boldsymbol{\alpha}\|_\infty\leqslant\|\boldsymbol{\alpha}\|_2\leqslant\sqrt{n}\|\boldsymbol{\alpha}\|_\infty$.

3. 设 $\|\cdot\|_a$ 与 $\|\cdot\|_b$ 是 $\mathbf{C}^n$ 上的两种向量范数，且 $k_1$，$k_2$ 是正常数，证明下列函数是 $\mathbf{C}^n$ 上的向量范数：

(1) $\max\{\|\boldsymbol{\alpha}\|_a, \|\boldsymbol{\alpha}\|_b\}$；

(2) $k_1\|\boldsymbol{\alpha}\|_a+k_2\|\boldsymbol{\alpha}\|_b$.

4. 设 $\|\cdot\|_m$ 是 $\mathbf{C}^{n\times n}$ 上的矩阵范数，$\boldsymbol{P}$ 是 $n$ 阶可逆矩阵，对任意的 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，定义 $\|\boldsymbol{A}\|=\|\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\|_m$，求证 $\|\cdot\|$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数.

5. 已知矩阵

$$\boldsymbol{A}=\begin{pmatrix}-1 & 0 & 2 & \mathrm{i}\\ 3 & 5 & -1 & 0\\ 1 & 2 & 0 & -1\\ 7 & -\mathrm{i} & 2 & -4\end{pmatrix},\quad \mathrm{i}^2=-1$$

计算 $\|\boldsymbol{A}\|_1$，$\|\boldsymbol{A}\|_\infty$. 如果 $\boldsymbol{x}=(-1, 2, 0, -\mathrm{i})^{\mathrm{T}}$，计算 $\|\boldsymbol{A}\boldsymbol{x}\|_1$，$\|\boldsymbol{A}\boldsymbol{x}\|_\infty$.

6. 已知

$$\boldsymbol{A}=\begin{pmatrix}1+\mathrm{i} & 0 & -3\\ 5 & 4\mathrm{i} & 0\\ -2 & 3 & 1\end{pmatrix}$$

试求 $\|\boldsymbol{A}\|_{m_1}$，$\|\boldsymbol{A}\|_F$，$\|\boldsymbol{A}\|_{m_\infty}$，$\|\boldsymbol{A}\|_1$，$\|\boldsymbol{A}\|_\infty$.

7. 设 $\boldsymbol{A}=(a_{ij})\in\mathbf{C}^{m\times n}$，定义

$$\|\boldsymbol{A}\|=\max\{m, n\}\cdot\max_{i,j}|a_{ij}|$$

则 $\|\boldsymbol{A}\|$ 是 $\mathbf{C}^{n\times n}$ 上的一种矩阵范数.

8. 设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，$\boldsymbol{\alpha}\in\mathbf{C}^n$，求证：矩阵范数 $\|\boldsymbol{A}\|_{m_1}$ 与向量范数 $p$-范数 $(1\leqslant p<+\infty)$ 相容.

9. 证明：矩阵范数 $\|\boldsymbol{A}\|_F$ 的 $\boldsymbol{U}$ 不变性($\boldsymbol{U}$ 为酉矩阵)，即

$$\| \boldsymbol{UA} \|_F = \| \boldsymbol{A} \|_F = \| \boldsymbol{AU} \|_F$$

10. 设

$$\boldsymbol{A} = \begin{bmatrix} 1 & 5 & -2 \\ -2 & 1 & 0 \\ 3 & -8 & 2 \end{bmatrix}$$

试求 $\mathrm{Cond}(\boldsymbol{A})_1$ 和 $\mathrm{Cond}(\boldsymbol{A})_\infty$.

11. 求下列方程组的极小范数最小二乘解：

$$\begin{cases} x_1 + 2x_2 + 3x_3 = 1 \\ x_1 + x_3 = 0 \\ 2x_1 + 2x_3 = 1 \\ 2x_1 + 4x_2 + 6x_3 = 3 \end{cases}$$

# 第6章 矩阵分析及其应用

前面几章中，主要研究了矩阵的代数运算. 本章将研究矩阵的分析运算，学习矩阵序列、矩阵级数、矩阵的微分与积分，以及上述这些知识的应用.

## 6.1 矩阵序列与矩阵级数

$m \times n$ 矩阵序列 $\boldsymbol{A}^{(1)}, \boldsymbol{A}^{(2)}, \cdots, \boldsymbol{A}^{(k)}, \cdots$，简记为 $\{\boldsymbol{A}^{(k)}\}$，其中

$$\boldsymbol{A}^{(k)} = \begin{pmatrix} a_{11}^{(k)} & a_{12}^{(k)} & \cdots & a_{1n}^{(k)} \\ a_{21}^{(k)} & a_{22}^{(k)} & \cdots & a_{2n}^{(k)} \\ \vdots & \vdots & & \vdots \\ a_{m1}^{(k)} & a_{m2}^{(k)} & \cdots & a_{mn}^{(k)} \end{pmatrix}, \quad k = 1, 2, \cdots$$

矩阵序列 $\{\boldsymbol{A}^{(k)}\}$ 中各矩阵的对应元素 $a_{ij}^{(k)}\ (i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n)$ 构成了 $m \times n$ 个数列.

**定义 6.1.1** 设有 $\mathbf{C}^{m\times n}$ 中矩阵序列 $\{\boldsymbol{A}^{(k)}\}$，$\boldsymbol{A}^{(k)} = (a_{ij}^{(k)})_{m\times n}$. 若

$$\lim_{k\to+\infty} a_{ij}^{(k)} = a_{ij} \quad (i = 1, 2, \cdots, m;\ j = 1, 2, \cdots, n)$$

则称矩阵序列 $\{\boldsymbol{A}^{(k)}\}$ 收敛于 $\boldsymbol{A} = (a_{ij})_{m\times n}$，称 $\boldsymbol{A}$ 为序列 $\{\boldsymbol{A}^{(k)}\}$ 的极限，记为

$$\boldsymbol{A}^{(k)} \to \boldsymbol{A}, \text{当 } k \to +\infty$$

或

$$\lim_{k\to+\infty} \boldsymbol{A}^{(k)} = \boldsymbol{A}$$

若矩阵序列不收敛，则称矩阵序列发散.

由定义可见，$\mathbf{C}^{m\times n}$ 中一个矩阵序列的收敛相当于 $m \times n$ 个数列同时收敛. 例如

$$A^{(k)}=\begin{pmatrix}\frac{(-1)^k}{k} & -1\\ \frac{1}{k} & \left(1+\frac{1}{k}\right)^k\end{pmatrix},\quad k=1,2,\cdots$$

$$\lim_{k\to+\infty}A^{(k)}=\lim_{k\to+\infty}\begin{pmatrix}\frac{(-1)^k}{k} & -1\\ \frac{1}{k} & \left(1+\frac{1}{k}\right)^k\end{pmatrix}=\begin{pmatrix}0 & -1\\ 0 & \mathrm{e}\end{pmatrix}=A$$

研究矩阵序列收敛性的常用且简单的方法,是利用矩阵的范数理论来研究矩阵序列的极限.

**定理 6.1.1**　设 $\|\cdot\|$ 是 $\mathbf{C}^{n\times n}$ 上任意一种矩阵范数,则 $\mathbf{C}^{n\times n}$ 中矩阵序列 $\{A^{(k)}\}$ 收敛于矩阵 $A$ 的充分必要条件是

$$\lim_{k\to+\infty}\|A^{(k)}-A\|=0$$

特别地,若 $A=0$,则 $\lim\limits_{k\to+\infty}A^{(k)}=0$ 的充分必要条件是

$$\lim_{k\to+\infty}\|A^{(k)}\|=0$$

**证**　由定义 6.1.1 知

$$\lim_{k\to+\infty}A^{(k)}=A\Leftrightarrow\lim_{k\to+\infty}a_{ij}^{(k)}=a_{ij}(i=1,2,\cdots,m;j=1,2,\cdots,n)$$

$$\Leftrightarrow\lim_{k\to+\infty}(a_{ij}^{(k)}-a_{ij})=0$$

$$\Leftrightarrow\lim_{k\to+\infty}\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}^{(k)}-a_{ij}|^2}=0$$

最后一式为

$$\lim_{k\to+\infty}\|A^{(k)}-A\|_F=0$$

由矩阵范数的等价性知,对于 $\mathbf{C}^{n\times n}$ 上任一矩阵范数 $\|\cdot\|$,必存在正常数 $C_1$,$C_2$,使

$$C_1\|A^{(k)}-A\|_F\leqslant\|A^{(k)}-A\|\leqslant C_2\|A^{(k)}-A\|_F$$

所以

$$\lim_{k\to+\infty}\|A^{(k)}-A\|_F=0\Leftrightarrow\lim_{k\to+\infty}\|A^{(k)}-A\|=0$$

故结论得证. □

特别地,若有 $\boldsymbol{A}=\boldsymbol{0}$, 则

$$\lim_{k\to+\infty}\boldsymbol{A}^{(k)}=\boldsymbol{0}\Leftrightarrow\lim_{k\to+\infty}\|\boldsymbol{A}^{(k)}-\boldsymbol{0}\|=\lim_{k\to+\infty}\|\boldsymbol{A}^{(k)}\|=0$$

由矩阵序列收敛性定义不难证明矩阵序列的极限运算满足下列定理.

**定理 6.1.2** 设 $\lim\limits_{k\to+\infty}\boldsymbol{A}^{(k)}=\boldsymbol{A}$, $\lim\limits_{k\to+\infty}\boldsymbol{B}^{(k)}=\boldsymbol{B}$, $a$、$b\in\mathbf{C}$, $\boldsymbol{A}^{(k)}$, $\boldsymbol{B}^{(k)}$, $\boldsymbol{A}$、$\boldsymbol{B}\in\mathbf{C}^{n\times n}$, 则

(i) $\lim\limits_{k\to+\infty}(a\boldsymbol{A}^{(k)}+b\boldsymbol{B}^{(k)})=a\boldsymbol{A}+b\boldsymbol{B}$;

(ii) $\lim\limits_{k\to+\infty}\boldsymbol{A}^{(k)}\boldsymbol{B}^{(k)}=\boldsymbol{A}\boldsymbol{B}$;

(iii) 当 $\boldsymbol{A}^{(k)}$ 与 $\boldsymbol{A}$ 均可逆时

$$\lim_{k\to+\infty}(\boldsymbol{A}^{(k)})^{-1}=\boldsymbol{A}^{-1}$$

注意定理 6.1.2(iii)中条件 $\boldsymbol{A}^{(k)}$, $\boldsymbol{A}$ 均可逆是不可缺少的. 考虑下面例子:

**例 6.1.1**

$$\boldsymbol{A}^{(k)}=\begin{pmatrix}1+\frac{1}{k} & -1\\ -1 & 1\end{pmatrix}$$

$$(\boldsymbol{A}^{(k)})^{-1}=\begin{pmatrix}k & k\\ k & k+1\end{pmatrix},\quad k=1,2,\cdots$$

$$\lim_{k\to+\infty}\boldsymbol{A}^{(k)}=\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix}=\boldsymbol{A}$$

但是 $\boldsymbol{A}$ 显然不可逆.

例 6.1.1 说明,虽然所有 $\boldsymbol{A}^{(k)}$ 可逆,且 $\lim\limits_{k\to+\infty}\boldsymbol{A}^{(k)}=\boldsymbol{A}$, 但不能保证 $\boldsymbol{A}$ 的可逆性.

在矩阵序列中,最常见的是由方阵的幂构成的矩阵序列.

**定义 6.1.2** 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$,若 $\lim\limits_{k\to+\infty}\boldsymbol{A}^{k}=\boldsymbol{0}$, 则称 $\boldsymbol{A}$ 为收敛矩阵.

**定理 6.1.3** 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$, 则 $\boldsymbol{A}$ 为收敛矩阵的充分必要条件是谱半径 $\rho(\boldsymbol{A})<1$.

**证** 设 $\boldsymbol{A}$ 的 Jordan 标准形为 $\boldsymbol{J}$,则存在可逆阵 $\boldsymbol{P}$,使

$$\boldsymbol{A}=\boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}$$

于是有

$$\boldsymbol{A}^{k}=\boldsymbol{P}\boldsymbol{J}^{k}\boldsymbol{P}^{-1}$$

可见, $\boldsymbol{A}^{k}\to\boldsymbol{0}$ 等价于 $\boldsymbol{J}^{k}\to\boldsymbol{0}$,又

$$\boldsymbol{J}^k = \mathrm{diag}\{\boldsymbol{J}_{r_1}^k(\lambda_1), \boldsymbol{J}_{r_2}^k(\lambda_2), \cdots, \boldsymbol{J}_{r_s}^k(\lambda_s)\}$$

其中 $\lambda_1, \lambda_2, \cdots, \lambda_s$ 为 $\boldsymbol{A}$ 的不同的特征值.

显然，$\boldsymbol{J}^k \to \boldsymbol{0}$ 等价于 $\boldsymbol{J}_{r_i}^k(\lambda_i) \to \boldsymbol{0}$ ($\forall i = 1, 2, \cdots, s$)，而

$$J_{r_i}^k(\lambda_i) = \begin{pmatrix} f_k(\lambda_i) & f_k'(\lambda_i) & \cdots & \dfrac{f_k^{(r_i-1)}(\lambda_i)}{(r_i-1)!} \\ & f_k(\lambda_i) & \cdots & \dfrac{f_k^{(r_i-2)}(\lambda_i)}{(r_i-2)!} \\ & & \ddots & \vdots \\ & & & f_k(\lambda_i) \end{pmatrix}$$

其中 $f_k(\lambda_i) = \lambda_i^k$. 所以，当 $|\lambda_i| < 1$ 时有

$$f_k^{(l)}(\lambda_i) \to 0, \quad l = 0, 1, 2, \cdots, r_i - 1$$

故

$$\boldsymbol{J}_{r_i}^k(\lambda_i) \to \boldsymbol{0}$$

从而

$$\boldsymbol{J}^k \to \boldsymbol{0}, \qquad \boldsymbol{A}^k \to \boldsymbol{0}$$

反之，若 $\boldsymbol{A}^k \to \boldsymbol{0}$，则

$$\boldsymbol{J}_{r_i}^{(k)}(\lambda_i) \to \boldsymbol{0}, \quad \forall i = 1, 2, \cdots, s$$

由此 $\lambda_i^k \to 0$，所以

$$|\lambda_i| < 1$$

故有

$$\rho(\boldsymbol{A}) < 1$$

□

**推论 6.1.1**　设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$，若 $\mathbf{C}^{n\times n}$ 上存在矩阵范数 $\|\cdot\|$，使 $\|\boldsymbol{A}\| < 1$，则 $\boldsymbol{A}$ 为收敛矩阵.

**推论 6.1.2**　方阵 $\boldsymbol{A}$ 的每个特征值 $\lambda$ 的模 $|\lambda|$ 都不大于矩阵 $\boldsymbol{A}$ 的任何一种范数 $\|\boldsymbol{A}\|$，即 $|\lambda| \leqslant \|\boldsymbol{A}\|$.

**证**　设 $\|\boldsymbol{A}\| = a$，作一个矩阵

$$\boldsymbol{B} = \frac{1}{a+\varepsilon}\boldsymbol{A}$$

其中 $\varepsilon$ 为任意正数，于是

$$\|\boldsymbol{B}\| = \frac{1}{a+\varepsilon}\|\boldsymbol{A}\| = \frac{a}{a+\varepsilon} < 1$$

由推论 6.1.1 知 $\boldsymbol{B}$ 为收敛矩阵，由定理 6.1.3 的证明知，$\boldsymbol{B}$ 的所有特征值的模都小于 1，而 $\boldsymbol{B}$ 的特征值就是 $\frac{\lambda}{a+\varepsilon}$，故

$$\left|\frac{\lambda}{a+\varepsilon}\right|<1$$

即

$$|\lambda|<a+\varepsilon$$

因为正数 $\varepsilon$ 可以任意小，因此

$$|\lambda|\leqslant a=\|\boldsymbol{A}\|$$ □

推论 6.1.2 反映了矩阵特征值与矩阵范数之间的一个基本关系.

推论 6.1.1 给出了判断矩阵 $\boldsymbol{A}$ 是否为收敛矩阵的一种方法. 即先验证是否有某一种矩阵范数满足 $\|\boldsymbol{A}\|<1$，如果不易找到这样的范数，则可以计算出 $\boldsymbol{A}$ 的所有特征值后，求 $\rho(\boldsymbol{A})$，再进一步判断.

**例 6.1.2** 判断下列矩阵是否为收敛矩阵：

$$\boldsymbol{A}=\begin{bmatrix}0.2 & 0.1 & 0.2\\0.5 & 0.4 & 0.4\\0.1 & 0.3 & 0.2\end{bmatrix}$$

**解** 由 $\|\boldsymbol{A}\|_1=0.8<1$，可知 $\boldsymbol{A}$ 为收敛矩阵.

应该注意到：存在矩阵的某种范数 $\|\cdot\|$，使得 $\|\boldsymbol{A}\|<1$ 仅是矩阵 $\boldsymbol{A}$ 为收敛矩阵的充分条件，不是必要条件. 例如

$$\boldsymbol{A}=\begin{pmatrix}0.9 & 0\\0.3 & 0.8\end{pmatrix}$$

由 $\rho(\boldsymbol{A})=0.9$ 知 $\boldsymbol{A}$ 为收敛矩阵，但是

$$\|\boldsymbol{A}\|_\infty=1.1,\qquad \|\boldsymbol{A}\|_1=1.2$$

$$\|\boldsymbol{A}\|_F\approx 1.24,\qquad \|\boldsymbol{A}\|_2\approx 1.02$$

**定义 6.1.3** 设 $\{\boldsymbol{A}^{(k)}\}$ 是 $\mathbf{C}^{m\times n}$ 上的矩阵序列，$\boldsymbol{A}^{(k)}=(a_{ij}^{(k)})\in\mathbf{C}^{m\times n}$，$\boldsymbol{S}=(s_{ij})_{m\times n}\in\mathbf{C}^{m\times n}$，无穷和

$$\boldsymbol{A}^{(0)}+\boldsymbol{A}^{(1)}+\cdots+\boldsymbol{A}^{(k)}+\cdots$$

称为矩阵级数，记为 $\sum_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$.

对任一正整数 $N$，称 $\boldsymbol{S}^{(N)}=\sum\limits_{k=0}^{N}\boldsymbol{A}^{(k)}$ 为矩阵级数 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 的部分和. 如果由部分和构成的矩阵序列 $\{\boldsymbol{S}^{(N)}\}$ 收敛于 $\boldsymbol{S}$，即 $\lim\limits_{N\to\infty}\boldsymbol{S}^{(N)}=\boldsymbol{S}$，则称矩阵级数 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 收敛，并称 $\boldsymbol{S}$ 为矩阵级数 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 的和，记为 $\boldsymbol{S}=\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$；不收敛的矩阵级数称为发散的.

$\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}=\boldsymbol{S}$ 等价于 $m\times n$ 个数项级数 $\sum\limits_{k=0}^{+\infty}a_{ij}^{(k)}=s_{ij}$.

**例 6.1.3**　已知

$$\boldsymbol{A}^{(k)}=\begin{bmatrix}\dfrac{1}{2^k} & \dfrac{\pi}{3\cdot 4^k}\\ 0 & \dfrac{1}{k(k+1)}\end{bmatrix},\quad k=1,2,\cdots$$

说明矩阵级数 $\sum\limits_{k=1}^{\infty}\boldsymbol{A}^{(k)}$ 的敛散性.

**解**　首先求矩阵级数的部分和

$$\boldsymbol{S}^{(N)}=\sum_{k=1}^{N}\boldsymbol{A}^{(k)}=\begin{bmatrix}1-\left(\dfrac{1}{2}\right)^N & \dfrac{\pi}{9}\left[1-\left(\dfrac{1}{4}\right)^N\right]\\ 0 & 1-\dfrac{1}{N+1}\end{bmatrix}$$

于是

$$\lim_{N\to+\infty}\boldsymbol{S}^{(N)}=\begin{bmatrix}1 & \dfrac{\pi}{9}\\ 0 & 1\end{bmatrix}$$

故矩阵级数 $\sum\limits_{k=1}^{+\infty}\boldsymbol{A}^{(k)}$ 收敛，且其和为

$$\boldsymbol{S}=\begin{bmatrix}1 & \dfrac{\pi}{9}\\ 0 & 1\end{bmatrix}$$

**定义 6.1.4**　设 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 是矩阵级数，其中 $\boldsymbol{A}^{(k)}=(a_{ij}^{(k)})\in\mathbf{C}^{m\times n}$. 如果 $m\times n$ 个数项级数 $\sum\limits_{k=0}^{+\infty}a_{ij}^{(k)}\,(i=1,2,\cdots,m;\ j=1,2,\cdots,n)$ 都绝对收敛，则称矩阵级数

$\sum_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 绝对收敛.

利用矩阵范数,可以将判定矩阵级数是否绝对收敛化为判定一个正项级数是否收敛的问题.

**定理 6.1.4** 设 $\boldsymbol{A}^{(k)} = (a_{ij}^{(k)}) \in \mathbf{C}^{m\times n}(k=0, 1, \cdots)$,则矩阵级数 $\sum_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 绝对收敛的充分必要条件是正项级数 $\sum_{k=0}^{+\infty} \| \boldsymbol{A}^{(k)} \|$ 收敛,其中 $\| \cdot \|$ 是 $\mathbf{C}^{m\times n}$上任一矩阵范数.

**证** 若 $\sum_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 绝对收敛,则 $m\times n$ 个级数 $\sum_{k=0}^{+\infty} | a_{ij}^{(k)} |$ 都收敛,故

$$\sum_{k=0}^{+\infty} | a_{ij}^{(k)} | < M_{ij}, \quad i=1, 2, \cdots, m; j=1, 2, \cdots, n$$

令 $M = \max\limits_{i, j} M_{ij}$,取矩阵范数 $\| \cdot \|$ 为 $m_1$ -范数,则有

$$\sum_{k=0}^{+\infty} \| \boldsymbol{A}^{(k)} \|_{m_1} = \sum_{k=0}^{+\infty} \left( \sum_{i=1}^{m} \sum_{j=1}^{n} | a_{ij}^{(k)} | \right) < mnM$$

这说明正项级数 $\sum_{k=0}^{+\infty} \| \boldsymbol{A}^{(k)} \|_{m_1}$ 收敛,由矩阵范数的等价性,对任意矩阵范数,正项级数 $\sum_{k=0}^{+\infty} \| \boldsymbol{A}^{(k)} \|$ 收敛.

若 $\sum_{k=0}^{+\infty} \| \boldsymbol{A}^{(k)} \|$ 收敛,由矩阵范数的等价性,所以 $\sum_{k=0}^{+\infty} \| \boldsymbol{A}^{(k)} \|_{m_1}$ 收敛. 又

$$| a_{ij}^{(k)} | \leqslant \sum_{i=1}^{m} \sum_{j=1}^{n} | a_{ij}^{(k)} | = \| \boldsymbol{A}^{(k)} \|_{m_1}, \quad i=1, 2, \cdots, m; j=1, 2, \cdots, n$$

由正项级数的比较判别法知 $\sum_{k=0}^{+\infty} | a_{ij}^{(k)} |$ 都收敛,即 $m\times n$ 个级数 $\sum_{k=0}^{+\infty} a_{ij}^{(k)} (i=1, 2, \cdots, m; j=1, 2, \cdots, n)$ 中的每一个级数都是绝对收敛的. 故 $\sum_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 绝对收敛. □

与数项级数类似,若矩阵级数 $\sum_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 绝对收敛,则 $\sum_{k=0}^{+\infty} \boldsymbol{A}^{(k)}$ 一定收敛,并且任意交换各项在和式中的次序所得的新的级数仍收敛,其和也不变.

下面介绍一类特殊的矩阵级数——矩阵幂级数.

**定义 6.1.5** 设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$, $C_k \in \mathbf{C}(k=0, 1, \cdots)$,称矩阵级数

$$\sum_{k=0}^{+\infty} C_k \boldsymbol{A}^k = C_0 \boldsymbol{E} + C_1 \boldsymbol{A} + C_2 \boldsymbol{A}^2 + \cdots + C_k \boldsymbol{A}^k + \cdots$$

为矩阵 $\boldsymbol{A}$ 的幂级数.

利用定义来判别矩阵幂级数的敛散性，需要判别 $n^2$ 个数项级数的敛散性. 当矩阵阶数 $n$ 较大时，通常是非常麻烦的事情. 显然，矩阵幂级数是复变量 $z$ 的幂级数 $\sum_{k=0}^{+\infty} C_k z^k$ 的推广. 我们知道，若幂级数 $\sum_{k=0}^{+\infty} C_k z^k$ 的收敛半径为 $R$，则对收敛圆 $|z| < R$ 内的所有 $z$，$\sum_{k=0}^{+\infty} C_k z^k$ 都是绝对收敛的. 因此，讨论矩阵幂级数 $\sum_{k=0}^{+\infty} C_k \boldsymbol{A}^k$ 的收敛性问题自然联想到复变量幂级数 $\sum_{k=0}^{+\infty} C_k z^k$ 的收敛半径.

**定理 6.1.5**　设幂级数 $\sum_{k=0}^{+\infty} C_k z^k$ 的收敛半径为 $R$，则

(i) 当 $\rho(\boldsymbol{A}) < R$ 时，矩阵幂级数 $\sum_{k=0}^{+\infty} C_k \boldsymbol{A}^k$ 绝对收敛；

(ii) 当 $\rho(\boldsymbol{A}) > R$ 时，矩阵幂级数 $\sum_{k=0}^{+\infty} C_k \boldsymbol{A}^k$ 发散.

**证**　(i) 设 $\rho(\boldsymbol{A}) < R$，则存在正数 $\varepsilon$，使 $\rho(\boldsymbol{A}) + \varepsilon < R$，由定理 5.3.3 知，存在 $\mathbf{C}^{n\times n}$ 上的矩阵范数 $\|\cdot\|_m$，使得

$$\|\boldsymbol{A}\|_m \leqslant \rho(\boldsymbol{A}) + \varepsilon < R$$

从而

$$\|C_k \boldsymbol{A}^k\|_m = |C_k| \, \|\boldsymbol{A}\|_m^k \leqslant |C_k| \, (\rho(\boldsymbol{A}) + \varepsilon)^k$$

因为幂级数

$$\sum_{k=0}^{+\infty} |C_k| \, (\rho(\boldsymbol{A}) + \varepsilon)^k$$

收敛，故矩阵幂级数 $\sum_{k=0}^{+\infty} C_k \boldsymbol{A}^k$ 绝对收敛.

(ii) 设 $\rho(\boldsymbol{A}) > R$，$\lambda_1, \lambda_2, \cdots, \lambda_n$ 为 $\boldsymbol{A}$ 的 $n$ 个特征值，则必存在一个特征值 $\lambda_t$，满足 $|\lambda_t| > R$. 由 Jordan 标准形理论，存在 $n$ 阶非奇异矩阵 $\boldsymbol{P}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \begin{pmatrix} \lambda_1 & k_1 & & & \\ & \lambda_2 & k_2 & & \\ & & \ddots & \ddots & \\ & & & \ddots & k_{n-1} \\ & & & & \lambda_n \end{pmatrix}$$

其中 $k_i$ 为 1 或 0，$i=1,2,\cdots,n-1$.

因为 $J^k$ 的对角元素为 $\lambda_i^k(i=1,2,\cdots,n)$，故矩阵幂级数 $\sum\limits_{k=0}^{+\infty}C_kJ^k$ 的对角元素为 $\sum\limits_{k=0}^{+\infty}C_k\lambda_i^k(i=1,2,\cdots,n)$. 其中 $\sum\limits_{k=0}^{+\infty}C_k\lambda_i^k$ 发散，所以 $\sum\limits_{k=0}^{+\infty}C_kJ^k$ 发散. 又 $A=PJP^{-1}$，故 $A^k=PJ^kP^{-1}$，$C_kA^k=P(C_kJ^k)P^{-1}$，从而可知 $\sum\limits_{k=0}^{+\infty}C_kA^k$ 发散. □

**推论 6.1.3** 若幂级数 $\sum\limits_{k=0}^{+\infty}C_kz^k$ 在整个平面上都收敛，则对任意的 $A\in\mathbf{C}^{n\times n}$，矩阵幂级数 $\sum\limits_{k=0}^{+\infty}C_kA^k$ 收敛.

**推论 6.1.4** 设幂级数 $\sum\limits_{k=0}^{+\infty}C_kz^k$ 的收敛半径为 $R$，$A\in\mathbf{C}^{n\times n}$，如果存在 $\mathbf{C}^{n\times n}$ 上的某一矩阵范数 $\|\cdot\|$，使得 $\|A\|<R$，则矩阵幂级数 $\sum\limits_{k=0}^{+\infty}C_kA^k$ 绝对收敛.

**例 6.1.4** 讨论矩阵幂级数

$$\sum_{k=1}^{\infty}\frac{1}{k^2}\begin{pmatrix}\frac{1}{6} & -\frac{1}{3}\\ -\frac{4}{3} & \frac{1}{6}\end{pmatrix}^k$$

的敛散性.

**解** 记

$$A=\begin{pmatrix}\frac{1}{6} & -\frac{1}{3}\\ -\frac{4}{3} & \frac{1}{6}\end{pmatrix}$$

求的 $A$ 的特征值为 $-\frac{1}{2},\frac{5}{6}$，则

$$\rho(A)=\frac{5}{6}$$

由于幂级数 $\sum\limits_{k=1}^{+\infty}\frac{1}{k^2}z^k$ 的收敛半径为 $R=1$，故由 $\rho(A)<R$ 知矩阵幂级数 $\sum\limits_{k=1}^{+\infty}\frac{1}{k^2}A^k$ 绝对收敛.

最后讨论一个特殊的矩阵幂级数.

**定义 6.1.6**　设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$，称 $\boldsymbol{A}$ 的幂级数

$$\sum_{k=0}^{+\infty} \boldsymbol{A}^k = \boldsymbol{E} + \boldsymbol{A} + \boldsymbol{A}^2 + \cdots + \boldsymbol{A}^k + \cdots$$

为 Neumann(诺伊曼)级数.

**定理 6.1.6**　Neumann 级数收敛的充分必要条件是 $\rho(\boldsymbol{A}) < 1$，且在收敛时，其和为 $(\boldsymbol{E}-\boldsymbol{A})^{-1}$.

**证**　设 $\sum\limits_{k=0}^{+\infty} \boldsymbol{A}^k$ 收敛，有 $\lim\limits_{k\to+\infty} \boldsymbol{A}^k = \boldsymbol{0}$，即 $\boldsymbol{A}$ 为收敛矩阵，从而有 $\rho(\boldsymbol{A}) < 1$.

设 $\rho(\boldsymbol{A}) < 1$，又幂级数 $\sum\limits_{k=0}^{+\infty} z^k$ 的收敛半径为 $R = 1$. 因此知 Neumann 级数绝对收敛.

当 $\sum\limits_{k=0}^{+\infty} \boldsymbol{A}^k$ 收敛时，$\rho(\boldsymbol{A}) < 1$，因此 $\boldsymbol{E}-\boldsymbol{A}$ 可逆，又因为

$$\begin{aligned} \boldsymbol{S}^{(N)}(\boldsymbol{E}-\boldsymbol{A}) &= (\boldsymbol{E}+\boldsymbol{A}+\cdots+\boldsymbol{A}^N)(\boldsymbol{E}-\boldsymbol{A}) \\ &= \boldsymbol{E}-\boldsymbol{A}^{N+1} \end{aligned}$$

故

$$\begin{aligned} \boldsymbol{S}^{(N)} &= (\boldsymbol{E}-\boldsymbol{A}^{N+1})(\boldsymbol{E}-\boldsymbol{A})^{-1} \\ &= (\boldsymbol{E}-\boldsymbol{A})^{-1} - \boldsymbol{A}^{N+1}(\boldsymbol{E}-\boldsymbol{A})^{-1} \end{aligned}$$

从而有

$$\boldsymbol{S} = \lim_{N\to+\infty} \boldsymbol{S}^{(N)} = (\boldsymbol{E}-\boldsymbol{A})^{-1}$$

□

**例 6.1.5**　设

$$\boldsymbol{A} = \begin{pmatrix} 0.2 & 0.5 \\ 0.7 & 0.4 \end{pmatrix}$$

试判断矩阵幂级数 $\sum\limits_{k=0}^{+\infty} \boldsymbol{A}^k$ 的敛散性，若收敛并求其和.

**解**　因 $\|\boldsymbol{A}\|_1 = 0.9 < 1$，故知

$$\rho(\boldsymbol{A}) \leqslant \|\boldsymbol{A}\|_1 < 1$$

所以 $\sum\limits_{k=0}^{+\infty} \boldsymbol{A}^k$ 收敛，且和为 $(\boldsymbol{E}-\boldsymbol{A})^{-1}$.

又

$$\boldsymbol{E}-\boldsymbol{A} = \begin{pmatrix} 0.8 & -0.5 \\ -0.7 & 0.6 \end{pmatrix} = \frac{1}{10}\begin{pmatrix} 8 & -5 \\ -7 & 6 \end{pmatrix}$$

$$(\boldsymbol{E}-\boldsymbol{A})^{-1}=\frac{10}{13}\begin{pmatrix}6 & 5\\ 7 & 8\end{pmatrix}$$

所以

$$\sum_{k=0}^{+\infty}\boldsymbol{A}^k=(\boldsymbol{E}-\boldsymbol{A})^{-1}=\frac{10}{13}\begin{pmatrix}6 & 5\\ 7 & 8\end{pmatrix}$$

## 6.2 矩阵函数及其计算

本节研究以矩阵为变量且取值为矩阵的函数,并称这类函数为矩阵函数.首先我们利用矩阵幂级数定义矩阵函数,然后再介绍计算矩阵函数的方法.

**定义 6.2.1** 设幂级数 $\sum_{k=0}^{+\infty}C_k z^k$ 的收敛半径为$R$,且当 $|z|<R$ 时,幂级数收敛于函数 $f(z)$,即

$$f(z)=\sum_{k=0}^{+\infty}C_k z^k,\quad |z|<R$$

如果 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 满足 $\rho(\boldsymbol{A})<R$,则称收敛的矩阵幂级数 $\sum_{k=0}^{+\infty}C_k\boldsymbol{A}^k$ 的和为矩阵函数,记为 $f(\boldsymbol{A})$,即

$$f(\boldsymbol{A})=\sum_{k=0}^{+\infty}C_k\boldsymbol{A}^k$$

由上述定义与高等数学及复变函数中关于幂级数展开式的知识,可以得到相应的矩阵函数.例如,对于如下函数的幂级数展开式:

$$\mathrm{e}^z=\sum_{k=0}^{+\infty}\frac{z^k}{k!},\qquad R=+\infty$$

$$\sin z=\sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k+1)!}z^{2k+1},\qquad R=+\infty$$

$$\cos z=\sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k)!}z^{2k},\qquad R=+\infty$$

$$(1-z)^{-1}=\sum_{k=0}^{+\infty}z^k,\qquad R=1$$

$$\ln(1+z)=\sum_{k=0}^{+\infty}\frac{(-1)^k}{k+1}z^{k+1},\qquad R=1$$

相应地，有矩阵函数

$$e^{A}=\sum_{k=0}^{+\infty}\frac{A^k}{k!},\qquad \forall A\in \mathbf{C}^{n\times n}$$

$$\sin A=\sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k+1)!}A^{2k+1},\qquad \forall A\in \mathbf{C}^{n\times n}$$

$$\cos A=\sum_{k=0}^{+\infty}\frac{(-1)^k}{(2k)!}A^{2k},\qquad \forall A\in \mathbf{C}^{n\times n}$$

$$(E-A)^{-1}=\sum_{k=0}^{+\infty}A^k,\qquad \rho(A)<1$$

$$\ln(E+A)=\sum_{k=0}^{+\infty}\frac{(-1)^k}{k+1}A^{k+1},\qquad \rho(A)<1$$

称 $e^A$ 为矩阵指数函数，$\sin A$ 为矩阵正弦函数，$\cos A$ 为矩阵余弦函数.

如果把矩阵函数 $f(A)$ 的变元 $A$ 换成 $At$，其中 $t$ 为参数，则相应地有

$$f(At)=\sum_{k=0}^{+\infty}C_k(At)^k$$

例如

$$e^{At}=\sum_{k=0}^{+\infty}\frac{1}{k!}(At)^k=\sum_{k=0}^{+\infty}\frac{t^k}{k!}A^k$$

类似地有 $\sin(At)$，$\cos(At)$ 等. 在实际应用中，经常需要求含参数的矩阵函数.

矩阵函数的计算问题是矩阵在应用中的关键问题. 由定义进行矩阵函数的计算是相当复杂的. 所以，研究如何方便地计算矩阵函数是非常有意义的，为此，我们介绍下列几种常用的算法.

**1. 递推公式法**

设 $f(\lambda)=|\lambda E-A|$，根据 Cayley-Hamilton 定理知，$f(A)=0$，由此找出矩阵方幂之间的关系，得出 $A$ 的递推关系式，从而计算给定的矩阵 $A$ 的函数. 举例说明如下.

**例 6.2.1** 已知 $A=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$，试求 $e^{At}$.

**解** 由 $f(\lambda)=|\lambda E-A|=\lambda^2+1$ 知 $A^2+E=0$，从而有

$$A^2=-E,\ A^3=-A,\ A^4=E,\ A^5=A,\ \cdots$$

即

$$A^{2k}=(-1)^kE,\qquad A^{2k+1}=(-1)^kA,\quad k=1,2,\cdots$$

所以

$$e^{\boldsymbol{A}t} = \sum_{k=0}^{+\infty} \frac{1}{k!}\boldsymbol{A}^k t^k = \left(1 - \frac{t^2}{2!} + \frac{t^4}{4!} - \cdots\right)\boldsymbol{E} + \left(t - \frac{t^3}{3!} + \frac{t^5}{5!} - \cdots\right)\boldsymbol{A}$$

$$= (\cos t)\boldsymbol{E} + (\sin t)\boldsymbol{A} = \begin{pmatrix} \cos t & \sin t \\ -\sin t & \cos t \end{pmatrix}$$

**2. 利用 Jordan 标准形计算**

设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$ 的 Jordan 标准形为 $\boldsymbol{J}$，即存在可逆矩阵 $\boldsymbol{P} \in \mathbf{C}^{n\times n}$，使得

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{J} = \mathrm{diag}\{\boldsymbol{J}_1, \boldsymbol{J}_2, \cdots, \boldsymbol{J}_i, \cdots, \boldsymbol{J}_s\}$$

其中

$$\boldsymbol{J}_i = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix}_{m_i}, \quad i = 1, 2, \cdots, s, \qquad \sum_{i=1}^{s} m_i = n$$

由此

$$\boldsymbol{A} = \boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1}, \qquad \boldsymbol{A}^k = \boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}$$

$$f(\boldsymbol{A}) = \sum_{k=0}^{+\infty} C_k \boldsymbol{A}^k = \boldsymbol{P}\left(\sum_{k=0}^{+\infty} C_k \boldsymbol{J}^k\right)\boldsymbol{P}^{-1}$$

$$= \boldsymbol{P}\begin{pmatrix} \sum_{k=0}^{+\infty} C_k \boldsymbol{J}_1^k & & & & \\ & \ddots & & & \\ & & \sum_{k=0}^{+\infty} C_k \boldsymbol{J}_i^k & & \\ & & & \ddots & \\ & & & & \sum_{k=0}^{+\infty} C_k \boldsymbol{J}_s^k \end{pmatrix}\boldsymbol{P}^{-1}$$

$$= \boldsymbol{P}\begin{pmatrix} f(\boldsymbol{J}_1) & & & & \\ & \ddots & & & \\ & & f(\boldsymbol{J}_i) & & \\ & & & \ddots & \\ & & & & f(\boldsymbol{J}_s) \end{pmatrix}\boldsymbol{P}^{-1}$$

$$= \boldsymbol{P}f(\boldsymbol{J})\boldsymbol{P}^{-1}$$

这里

$$f(\boldsymbol{J}_i)=\sum_{k=0}^{+\infty}C_k\boldsymbol{J}_i^k$$

$$=\begin{pmatrix}\sum_{k=0}^{+\infty}C_k\lambda_i^k & \frac{1}{1!}\sum_{k=0}^{+\infty}C_kC_k^1\lambda_i^{k-1} & \cdots & \frac{1}{(m_i-1)!}\sum_{k=0}^{+\infty}C_kC_k^{m_i-1}\lambda_i^{k-m_i+1}\\ & \sum_{k=0}^{+\infty}C_k\lambda_i^k & \ddots & \vdots\\ & & \ddots & \frac{1}{1!}\sum_{k=0}^{+\infty}C_kC_k^1\lambda_i^{k-1}\\ & & & \sum_{k=0}^{+\infty}C_k\lambda_i^k\end{pmatrix}$$

$$=\begin{pmatrix}f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(m_i-1)!}f^{(m_i-1)}(\lambda_i)\\ & f(\lambda_i) & \ddots & \vdots\\ & & \ddots & \frac{1}{1!}f'(\lambda_i)\\ & & & f(\lambda_i)\end{pmatrix}$$

特别地，如果矩阵 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 相似于对角矩阵 $\boldsymbol{\Lambda}$，即存在可逆矩阵 $\boldsymbol{P}$，使得

$$\boldsymbol{A}=\boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}^{-1},\quad \boldsymbol{\Lambda}=\operatorname{diag}(\lambda_1,\lambda_2,\cdots,\lambda_n)$$

于是

$$f(\boldsymbol{A})=\boldsymbol{P}\begin{pmatrix}f(\lambda_1) & & & \\ & f(\lambda_2) & & \\ & & \ddots & \\ & & & f(\lambda_n)\end{pmatrix}\boldsymbol{P}^{-1}$$

**例 6.2.2** 已知

$$\boldsymbol{A}=\begin{pmatrix}-1 & 0 & 1\\ 1 & 2 & 0\\ -4 & 0 & 3\end{pmatrix}$$

试求 $e^{\boldsymbol{A}}$，$\sin\boldsymbol{A}t$.

**解** 可以求得 $\boldsymbol{A}$ 的 Jordan 标准形

$$J=\begin{pmatrix}1&1&0\\0&1&0\\0&0&2\end{pmatrix}$$

且有

$$P=\begin{pmatrix}1&0&0\\-1&-1&1\\2&1&0\end{pmatrix},\qquad P^{-1}AP=J$$

当 $f(z)=e^z$ 时，$f(1)=e$，$f'(1)=e$，于是有

$$e^A=P\begin{pmatrix}e&e&0\\0&e&0\\0&0&e^2\end{pmatrix}P^{-1}$$

$$=\begin{pmatrix}-e&0&e\\3e-e^2&e^2&-2e+e^2\\-4e&0&3e\end{pmatrix}$$

当 $f(z)=\sin zt$ 时

$$f(1)=\sin t,\qquad f'(1)=t\cos t$$

于是有

$$\sin At=P\begin{pmatrix}\sin t&t\cos t&0\\0&\sin t&0\\0&0&\sin 2t\end{pmatrix}P^{-1}$$

$$=\begin{pmatrix}\sin t-2t\cos t&0&t\cos t\\\sin t+2t\cos t-\sin 2t&\sin 2t&-t\cos t-\sin t+\sin 2t\\-4t\cos t&0&2t\cos t+\sin t\end{pmatrix}$$

用 Jordan 标准形的方法来求矩阵函数的难点在于要求 $J$-矩阵及可逆矩阵 $P$、$P^{-1}$，所以即使 $A$ 形式上比较简单，这一工作量也较繁杂. 为避开这一点，我们有下列方法.

**3. 最小多项式法(待定系数法)**

设矩阵 $A\in C^{n\times n}$，$A$ 的最小多项式为

$$\varphi(\lambda)=(\lambda-\lambda_1)^{k_1}(\lambda-\lambda_2)^{k_2}\cdots(\lambda-\lambda_s)^{k_s}$$

其中，$\lambda_1$，$\lambda_2$，…，$\lambda_s$ 是 $A$ 的全部互异特征值；$k_1+k_2+\cdots+k_s=m$.

设矩阵函数 $f(\boldsymbol{A})=\sum_{k=0}^{+\infty}C_k\boldsymbol{A}^k$，$f(\lambda)=\sum_{k=0}^{+\infty}C_k\lambda^k$，由带余除法

$$f(\lambda)=q(\lambda)\varphi(\lambda)+r(\lambda) \tag{6.2.1}$$

其中 $r(\lambda)$是次数低于 $m$ 的$\lambda$ 的多项式，记为

$$r(\lambda)=a_{m-1}\lambda^{m-1}+\cdots+a_1\lambda+a_0$$

由于 $\varphi(\boldsymbol{A})=\boldsymbol{0}$，故有

$$f(\boldsymbol{A})=r(\boldsymbol{A})=a_{m-1}\boldsymbol{A}^{m-1}+\cdots+a_1\boldsymbol{A}+a_0\boldsymbol{E} \tag{6.2.2}$$

又 $\varphi^{(j)}(\lambda_i)=0$ $(j=0,1,\cdots,k_i-1;\ i=1,2,\cdots,m)$，因此对式(6.2.1)两边关于$\lambda$ 求导得

$$\left.\frac{\mathrm{d}^j}{\mathrm{d}\lambda^j}f(\lambda)\right|_{\lambda=\lambda_i}=\left.\frac{\mathrm{d}^j}{\mathrm{d}\lambda^j}r(\lambda)\right|_{\lambda=\lambda_i} \tag{6.2.3}$$

由此，可以得到以 $a_{m-1}$，$\cdots$，$a_1$，$a_0$ 为未知变量的线性方程组，求解即得到待定的系数 $a_{m-1}$，$\cdots$，$a_1$，$a_0$，从而由(6.2.2)得到矩阵函数 $f(\boldsymbol{A})$.

如果要计算矩阵函数 $f(\boldsymbol{A}t)=\sum_{k=0}^{\infty}C_k\boldsymbol{A}^kt^k$，也可以用类似的待定系数的方法，只需要将上面的函数 $f(\lambda)$用 $f(\lambda t)$代替. 只需注意 $t$ 是一个参数，式(6.2.3)中对函数 $f(\lambda t)$关于$\lambda$ 求导时 $t$ 看成一个常数.

**例 6.2.3** 设

$$\boldsymbol{A}=\begin{pmatrix}-1&-2&6\\-1&0&3\\-1&-1&4\end{pmatrix}$$

试求 $\mathrm{e}^{\boldsymbol{A}}$.

**解** 先求矩阵 $\boldsymbol{A}$ 的最小多项式. 由 $|\lambda\boldsymbol{E}-\boldsymbol{A}|=(\lambda-1)^3$ 得特征值

$$\lambda=1$$

又

$$\boldsymbol{A}-\boldsymbol{E}\neq\boldsymbol{0},\qquad(\boldsymbol{A}-\boldsymbol{E})^2=\boldsymbol{0}$$

所以 $\boldsymbol{A}$ 的最小多项式为

$$\varphi(\lambda)=(\lambda-1)^2$$

其次，设 $r(\lambda)=a_0+a_1\lambda$，$f(\lambda)=\mathrm{e}^{\lambda}$，从而

$$r(1) = a_0 + a_1 = \mathrm{e} = f(1)$$

$$r'(1) = a_1 = \mathrm{e} = f'(1)$$

解之,得

$$a_1 = \mathrm{e}, \quad a_0 = 0$$

从而

$$\mathrm{e}^{\boldsymbol{A}} = a_0\boldsymbol{E} + a_1\boldsymbol{A} = \begin{bmatrix} -\mathrm{e} & -2\mathrm{e} & 6\mathrm{e} \\ -\mathrm{e} & 0 & 3\mathrm{e} \\ -\mathrm{e} & -\mathrm{e} & 4\mathrm{e} \end{bmatrix}$$

**例 6.2.4** 设

$$\boldsymbol{A} = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 2 & 2 \\ -4 & 0 & 3 \end{bmatrix}$$

试求 $\mathrm{e}^{\boldsymbol{A}t}$, $\mathrm{e}^{\boldsymbol{A}}$, $\cos\boldsymbol{A}$.

**解** 先求矩阵 $\boldsymbol{A}$ 的最小多项式. 由 $|\lambda\boldsymbol{E} - \boldsymbol{A}| = (\lambda - 1)^2(\lambda - 2)$ 得特征值

$$\lambda_1 = 1, \qquad \lambda_2 = 2$$

由于 $(\boldsymbol{A} - \boldsymbol{E})(\boldsymbol{A} - 2\boldsymbol{E}) \neq \boldsymbol{0}$, 所以 $\boldsymbol{A}$ 的最小多项式即为

$$\varphi(\lambda) = |\lambda\boldsymbol{E} - \boldsymbol{A}| = (\lambda - 1)^2(\lambda - 2)$$

设 $r(\lambda) = a_0 + a_1\lambda + a_2\lambda^2$, $f(\lambda) = \mathrm{e}^{\lambda t}$, 从而

$$r(1) = a_0 + a_1 + a_2 = \mathrm{e}^t = f(1)$$

$$r'(1) = a_1 + 2a_2 = t\mathrm{e}^t = f'(1)$$

$$r(2) = a_0 + 2a_1 + 4a_2 = \mathrm{e}^{2t} = f(2)$$

解之,得

$$a_0 = \mathrm{e}^{2t} - 2t\mathrm{e}^t$$

$$a_1 = -2\mathrm{e}^{2t} + 2\mathrm{e}^t + 3t\mathrm{e}^t$$

$$a_2 = \mathrm{e}^{2t} - \mathrm{e}^t - t\mathrm{e}^t$$

从而

$$e^{At}=a_0E+a_1A+a_2A^2=\begin{pmatrix} e^t-2te^t & 0 & te^t \\ -e^{2t}+e^t+2te^t & e^{2t} & e^{2t}-e^t-te^t \\ -4te^t & 0 & 2te^t+e^t \end{pmatrix}$$

其次，令 $t=1$，得

$$e^A=\begin{pmatrix} -e & 0 & e \\ -e^2+3e & e^2 & e^2-2e \\ -4e & 0 & 3e \end{pmatrix}$$

最后，令 $f(\lambda)=\cos\lambda$，由

$$r(1)=a_0+a_1+a_2=\cos 1=f(1)$$

$$r'(1)=a_1+2a_2=-\sin 1=f'(1)$$

$$r(2)=a_0+2a_1+4a_2=\cos 2=f(2)$$

得

$$a_0=2\sin 1+\cos 2$$

$$a_1=-3\sin 1+2\cos 1-2\cos 2$$

$$a_2=\sin 1-\cos 1+\cos 2$$

从而

$$\begin{aligned}\cos A&=a_0E+a_1A+a_2A^2\\ &=\begin{pmatrix} 2\sin 1+\cos 2 & 0 & -\sin 1 \\ -2\sin 1+\cos 1-\cos 2 & \cos 2 & \sin 1-\cos 1+\cos 2 \\ 4\sin 1 & 0 & -2\sin 1+\cos 1 \end{pmatrix}\end{aligned}$$

由上述方法及例子可知矩阵函数的另一定义.

**定义 6.2.2**　设 $A\in C^{n\times n}$，$\lambda_i$ 为 $A$ 的互异特征值，$i=1,2,\cdots,s$，$J$ 为 $A$ 的 Jordan 标准形，即存在可逆矩阵 $P\in C^{n\times n}$，使 $P^{-1}AP=J$.

$$J=\mathrm{diag}\{J_1,J_2,\cdots,J_s\}$$

$$J_i=\begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ & & & & \lambda_i \end{pmatrix}_{r_i}$$

$f(z)$在 $\lambda_i$ 处有直到 $r_i-1$ 阶导数，$f(\boldsymbol{A})$定义为

$$f(\boldsymbol{A})=\boldsymbol{P}\begin{bmatrix} f(\boldsymbol{J}_1) & & & & \\ & \ddots & & & \\ & & f(\boldsymbol{J}_i) & & \\ & & & \ddots & \\ & & & & f(\boldsymbol{J}_s) \end{bmatrix}\boldsymbol{P}^{-1}$$

其中

$$f(\boldsymbol{J}_i)=\begin{bmatrix} f(\lambda_i) & \frac{1}{1!}f'(\lambda_i) & \cdots & \frac{1}{(r_i-1)!}f^{(r_i-1)}(\lambda_i) \\ & f(\lambda_i) & \ddots & \vdots \\ & & \ddots & \frac{1}{1!}f'(\lambda_i) \\ & & & f(\lambda_i) \end{bmatrix}$$

这一定义对于那些不能展开成收敛的幂级数的函数也能定义出矩阵函数.

**例 6.2.5** 设

$$\boldsymbol{A}=\begin{bmatrix} 5 & -4 & 4 \\ 1 & 0 & 5 \\ 1 & -1 & 6 \end{bmatrix}$$

试求 $\ln\boldsymbol{A}$.

**解**

$$\boldsymbol{A}=\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 5 & 1 \\ 0 & 0 & 5 \end{bmatrix}\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}^{-1}$$

故

$$\ln\boldsymbol{A}=\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 & 0 \\ 0 & \ln 5 & \frac{1}{5} \\ 0 & 0 & \ln 5 \end{bmatrix}\begin{bmatrix} 0 & 1 & -1 \\ 0 & 0 & 1 \\ 1 & -1 & 1 \end{bmatrix}$$

$$=\begin{bmatrix} \ln 5 & -\ln 5 & \ln 5 \\ \frac{1}{5} & -\frac{1}{5} & \frac{1}{5}+\ln 5 \\ \frac{1}{5} & -\frac{1}{5} & \frac{1}{5}+\ln 5 \end{bmatrix}$$

## 6.3 矩阵的微分与积分

矩阵函数与函数矩阵的微分、积分在研究微分方程组及优化问题中是非常重要的.

**定义 6.3.1** 以变量 $t$ 的函数为元素的矩阵 $\boldsymbol{A}(t)=(a_{ij}(t))_{m\times n}$ 称为函数矩阵,其中 $a_{ij}(t)(i=1,2,\cdots,m;j=1,2,\cdots,n)$ 都是变量 $t$ 的函数. 若 $t\in[a,b]$,则称 $\boldsymbol{A}(t)$ 是定义在$[a,b]$上的函数;又若每个 $a_{ij}(t)$ 在$[a,b]$上连续、可微、可积,则称 $\boldsymbol{A}(t)$ 在$[a,b]$上连续,可微、可积. 当 $\boldsymbol{A}(t)$ 可微时,规定其导数为

$$\boldsymbol{A}'(t)=(a'_{ij}(t))_{m\times n} \quad 或 \quad \frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)=\left(\frac{\mathrm{d}}{\mathrm{d}t}a_{ij}(t)\right)_{m\times n} \tag{6.3.1}$$

当 $\boldsymbol{A}(t)$ 在$[a,b]$上可积时,规定 $\boldsymbol{A}(t)$ 在$[a,b]$上的积分为

$$\int_a^b \boldsymbol{A}(t)\mathrm{d}t=\left(\int_a^b a_{ij}(t)\mathrm{d}t\right)_{m\times n} \tag{6.3.2}$$

不难证明函数矩阵的微分与积分具有下列定理.

**定理 6.3.1** 设 $\boldsymbol{A}(t)$ 与 $\boldsymbol{B}(t)$ 都是可微的,则

(i) $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t)+\boldsymbol{B}(t))=\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)+\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)$;

(ii) 当 $k(t)$ 为可微函数时,有

$$\frac{\mathrm{d}}{\mathrm{d}t}(k(t)\boldsymbol{A}(t))=\left(\frac{\mathrm{d}}{\mathrm{d}t}k(t)\right)\boldsymbol{A}(t)+k(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$$

(iii) $\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}(t)\boldsymbol{B}(t))=\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right)\boldsymbol{B}(t)+\boldsymbol{A}(t)\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{B}(t)$;

(iv) 当 $u=f(t)$ 关于 $t$ 可微时,有

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(u)=f'(t)\frac{\mathrm{d}}{\mathrm{d}u}\boldsymbol{A}(u)$$

(v) 当 $\boldsymbol{A}^{-1}(t)$ 是可微函数矩阵时,有

$$\frac{\mathrm{d}}{\mathrm{d}t}(\boldsymbol{A}^{-1}(t))=-\boldsymbol{A}^{-1}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right)\boldsymbol{A}^{-1}(t)$$

**证** 只证(v).

因为 $\boldsymbol{A}(t)\boldsymbol{A}^{-1}(t)=\boldsymbol{E}$,两边求导得

$$A(t)\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}(t)+\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)A^{-1}(t)=\mathbf{0}$$

从而

$$\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}(t)=-A^{-1}(t)\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)A^{-1}(t)$$ □

**定理 6.3.2** 设 $\boldsymbol{A}(t)$，$\boldsymbol{B}(t)$是区间$[a, b]$上的可积函数矩阵，$\boldsymbol{A}$、$\boldsymbol{B}$ 是常数矩阵，$\lambda\in\mathbf{C}$，则

(i) $\int_a^b(\boldsymbol{A}(t)+\boldsymbol{B}(t))\mathrm{d}t=\int_a^b\boldsymbol{A}(t)\mathrm{d}t+\int_a^b\boldsymbol{B}(t)\mathrm{d}t$；

(ii) $\int_a^b\lambda\boldsymbol{A}(t)\mathrm{d}t=\lambda\int_a^b\boldsymbol{A}(t)\mathrm{d}t$；

(iii) $\int_a^b\boldsymbol{A}(t)\boldsymbol{B}\mathrm{d}t=\left(\int_a^b\boldsymbol{A}(t)\mathrm{d}t\right)\boldsymbol{B}$，$\int_a^b\boldsymbol{A}\boldsymbol{B}(t)\mathrm{d}t=\boldsymbol{A}\left(\int_a^b\boldsymbol{B}(t)\mathrm{d}t\right)$；

(iv) 当 $\boldsymbol{A}(t)$在$[a, b]$上连续时，对任意 $t\in(a, b)$，有

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\int_a^t\boldsymbol{A}(\tau)\mathrm{d}(\tau)\right)=\boldsymbol{A}(t)$$

(v) 当 $\boldsymbol{A}(t)$在$[a, b]$上连续可微时，有

$$\int_a^b\boldsymbol{A}'(t)\mathrm{d}t=\boldsymbol{A}(b)-\boldsymbol{A}(a)$$

以上介绍了函数矩阵的微积分的一些运算法则，由于$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$仍是函数矩阵，若$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$仍为可导矩阵，则可以定义其二阶导数. 不难给出函数矩阵的高阶导数

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\boldsymbol{A}(t)=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}^{k-1}}{\mathrm{d}t^{k-1}}\boldsymbol{A}(t)\right)\tag{6.3.3}$$

**例 6.3.1** 已知 $\boldsymbol{A}(t)=\begin{pmatrix}\cos t & \sin t\\ -\sin t & \cos t\end{pmatrix}$，求：

(1) $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)$，$\left|\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right|$，$\frac{\mathrm{d}}{\mathrm{d}t}|\boldsymbol{A}(t)|$，$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}^{-1}(t)$，$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\boldsymbol{A}(t)$；

(2) $\int_0^{\frac{\pi}{2}}\boldsymbol{A}(t)\mathrm{d}t$，$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^{t^2}\boldsymbol{A}(s)\mathrm{d}s$.

**解** (1) $\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)=\begin{pmatrix}-\sin t & \cos t\\ -\cos t & -\sin t\end{pmatrix}$；  $\left|\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}(t)\right|=1$

由 $|A(t)|=1$ 知

$$\frac{\mathrm{d}}{\mathrm{d}t}|A(t)|=0$$

由 $A^{-1}(t)=\begin{pmatrix}\cos t & -\sin t\\ \sin t & \cos t\end{pmatrix}$ 知

$$\frac{\mathrm{d}}{\mathrm{d}t}A^{-1}(t)=\begin{pmatrix}-\sin t & -\cos t\\ \cos t & -\sin t\end{pmatrix}$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}A(t)=\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}}{\mathrm{d}t}A(t)\right)=\begin{pmatrix}-\cos t & -\sin t\\ \sin t & -\cos t\end{pmatrix}$$

(2)
$$\int_0^{\frac{\pi}{2}}A(t)\mathrm{d}t=\begin{pmatrix}1 & 1\\ -1 & 1\end{pmatrix}$$

利用变上限函数的导数公式，有

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_0^{t^2}A(s)\mathrm{d}s=2tA(t^2)=2t\begin{bmatrix}\cos t^2 & \sin t^2\\ -\sin t^2 & \cos t^2\end{bmatrix}$$

**例 6.3.2**　不论 $A$ 为任何常量方阵，都有

(i) $\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At}=A\mathrm{e}^{At}=\mathrm{e}^{At}A$；

(ii) $\frac{\mathrm{d}}{\mathrm{d}t}\cos At=-A(\sin At)=-(\sin At)A$；

(iii) $\frac{\mathrm{d}}{\mathrm{d}t}\sin At=A(\cos At)=(\cos At)A$.

**证**　(1) 由 $\mathrm{e}^{At}$ 的定义及矩阵级数理论易知

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{At}&=\frac{\mathrm{d}}{\mathrm{d}t}\left(E+\frac{1}{1!}At+\frac{1}{2!}A^2t^2+\cdots\right)\\&=A+\frac{1}{1!}A^2t+\frac{1}{2!}A^3t^2+\cdots\\&=A\left(E+\frac{1}{1!}At+\frac{1}{2!}A^2t^2+\cdots\right)\\&=A\mathrm{e}^{At}\end{aligned}$$

又由 $A^{m+1}t^m=(A^mt^m)A$，故有

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{\boldsymbol{A}t}=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{A}$$

类似地可证明(ii)、(iii).

**例 6.3.3** 设 $\boldsymbol{A}=(a_{ij}(t))_{n\times n}$，试证：

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\boldsymbol{A}=\mathrm{tr}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}$$

即说明 tr 与$\frac{\mathrm{d}}{\mathrm{d}t}$两种运算对于方阵来说可以交换次序.

**证** 因为

$$\mathrm{tr}\boldsymbol{A}=a_{11}(t)+a_{22}(t)+\cdots+a_{nn}(t)$$

所以

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\boldsymbol{A}=\frac{\mathrm{d}}{\mathrm{d}t}a_{11}(t)+\frac{\mathrm{d}}{\mathrm{d}t}a_{22}(t)+\cdots+\frac{\mathrm{d}}{\mathrm{d}t}a_{nn}(t)$$

又由于

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}=\begin{pmatrix}\frac{\mathrm{d}}{\mathrm{d}t}a_{11}(t) & \frac{\mathrm{d}}{\mathrm{d}t}a_{12}(t) & \cdots & \frac{\mathrm{d}}{\mathrm{d}t}a_{1n}(t)\\ \frac{\mathrm{d}}{\mathrm{d}t}a_{21}(t) & \frac{\mathrm{d}}{\mathrm{d}t}a_{22}(t) & \cdots & \frac{\mathrm{d}}{\mathrm{d}t}a_{2n}(t)\\ \vdots & \vdots & & \vdots\\ \frac{\mathrm{d}}{\mathrm{d}t}a_{n1}(t) & \frac{\mathrm{d}}{\mathrm{d}t}a_{n2}(t) & \cdots & \frac{\mathrm{d}}{\mathrm{d}t}a_{nn}(t)\end{pmatrix}$$

故

$$\mathrm{tr}\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{A}\right)=\frac{\mathrm{d}}{\mathrm{d}t}a_{11}(t)+\frac{\mathrm{d}}{\mathrm{d}t}a_{22}(t)+\cdots+\frac{\mathrm{d}}{\mathrm{d}t}a_{nn}(t)=\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{tr}\boldsymbol{A}$$

在自动控制理论及一些实际问题中，还经常遇到矩阵特殊的导数问题. 现介绍如下.

**定义 6.3.2** 设 $\boldsymbol{A}\in\mathbf{C}^{m\times n}$，$f(\boldsymbol{A})$是矩阵 $\boldsymbol{A}$ 的数量值函数(即自变量为 $\boldsymbol{A}$，函数值为数量 $f(\boldsymbol{A})$的 $m\times n$ 元函数). 若 $f(\boldsymbol{A})$关于 $\boldsymbol{A}$ 的任一元素 $a_{ij}$ $(i=1, 2, \cdots, m;$ $j=1, 2, \cdots, n)$ 的偏导数 $\frac{\partial f}{\partial a_{ij}}$都存在，则称

$$\left(\frac{\partial f}{\partial a_{ij}}\right)_{m\times n}=\begin{pmatrix}\frac{\partial f}{\partial a_{11}} & \frac{\partial f}{\partial a_{12}} & \cdots & \frac{\partial f}{\partial a_{1n}}\\ \frac{\partial f}{\partial a_{21}} & \frac{\partial f}{\partial a_{22}} & \cdots & \frac{\partial f}{\partial a_{2n}}\\ \vdots & \vdots & & \vdots\\ \frac{\partial f}{\partial a_{m1}} & \frac{\partial f}{\partial a_{m2}} & \cdots & \frac{\partial f}{\partial a_{mn}}\end{pmatrix}$$

为 $f(\boldsymbol{A})$关于 $\boldsymbol{A}=(a_{ij})_{m\times n}$ 的导数，记为 $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{A}}f(\boldsymbol{A})\left(或\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{A}}\right)$.

当 $m=n=1$ 时，表示的是一般的一元函数的导数，当 $m$ 或 $n$ 等于 1 时，就是 $n$ 元($m$ 元)函数的梯度的推广.

**例 6.3.4**　设

$$\boldsymbol{A}=\begin{pmatrix}a & b & c\\ d & e & f\end{pmatrix}$$

$$F(\boldsymbol{A})=a^2+b^2+c^2+d^2-2e+15f$$

试求 $\frac{\mathrm{d}F}{\mathrm{d}\boldsymbol{A}}$.

**解**

$$\frac{\mathrm{d}F}{\mathrm{d}\boldsymbol{A}}=\begin{pmatrix}\frac{\partial F}{\partial a} & \frac{\partial F}{\partial b} & \frac{\partial F}{\partial c}\\ \frac{\partial F}{\partial d} & \frac{\partial F}{\partial e} & \frac{\partial F}{\partial f}\end{pmatrix}=\begin{pmatrix}2a & 2b & 2c\\ 2d & -2 & 15\end{pmatrix}$$

**例 6.3.5**　设

$$\boldsymbol{A}=\begin{pmatrix}x_{11} & x_{12}\\ x_{21} & x_{22}\end{pmatrix}$$

$$f(\boldsymbol{A})=\mathrm{tr}\boldsymbol{A}=x_{11}+x_{22}$$

试求 $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{A}}$.

**解**

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{A}}=\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{A}}\mathrm{tr}\boldsymbol{A}=\begin{pmatrix}\frac{\partial f}{\partial x_{11}} & \frac{\partial f}{\partial x_{12}}\\ \frac{\partial f}{\partial x_{21}} & \frac{\partial f}{\partial x_{22}}\end{pmatrix}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

推广知，当 $\boldsymbol{A}=(a_{ij})_n \in \mathbf{R}^{n\times n}$ 时

$$f(\boldsymbol{A})=\operatorname{tr}\boldsymbol{A}$$

则有

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{A}}=\boldsymbol{E}_n$$

**例 6.3.6** 设

$$\boldsymbol{A}=(a_1, a_2, \cdots, a_n)^{\mathrm{T}}$$

$$f(\boldsymbol{A})=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}=a_1^2+a_2^2+\cdots+a_n^2$$

试求：(1) $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{A}}$；(2) $\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{A}^{\mathrm{T}}}$.

**解** (1)

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{A}}=\left(\frac{\partial f}{\partial a_1}, \frac{\partial f}{\partial a_2}, \cdots, \frac{\partial f}{\partial a_n}\right)^{\mathrm{T}}=(2a_1, 2a_2, \cdots, 2a_n)^{\mathrm{T}}=2\boldsymbol{A}$$

(2)

$$\frac{\partial f}{\partial \boldsymbol{A}^{\mathrm{T}}}=\left(\frac{\partial f}{\partial a_1}, \frac{\partial f}{\partial a_2}, \cdots, \frac{\partial f}{\partial a_n}\right)=(2a_1, 2a_2, \cdots, 2a_n)=2\boldsymbol{A}^{\mathrm{T}}$$

**例 6.3.7** 设 $\boldsymbol{A}=(a_{ij})_n$ 是给定的矩阵，$\boldsymbol{\alpha}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}$ 是向量变量，且 $f(\boldsymbol{\alpha})=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\alpha}$，试求$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{\alpha}}$.

**解**

$$f(\boldsymbol{\alpha})=\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{\alpha}=\sum_{s=1}^{n}\sum_{k=1}^{n}x_s a_{sk} x_k=\sum_{s=1}^{n}x_s\left(\sum_{k=1}^{n}a_{sk}x_k\right)$$

而

$$\frac{\partial f}{\partial x_j}=x_1a_{1j}+\cdots+x_{j-1}a_{j-1j}+\left(\sum_{k=1}^{n}a_{jk}x_k+x_ja_{jj}\right)+x_{j+1}a_{j+1j}+\cdots+x_na_{nj}$$

$$=\sum_{s=1}^{n}a_{sj}x_s+\sum_{k=1}^{n}a_{jk}x_k$$

所以

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{\alpha}}=\begin{pmatrix}\frac{\partial f}{\partial x_1}\\ \vdots\\ \frac{\partial f}{\partial x_n}\end{pmatrix}=\begin{pmatrix}\sum_{s=1}^{n}a_{s1}x_s+\sum_{k=1}^{n}a_{1k}x_k\\ \vdots\\ \sum_{s=1}^{n}a_{sn}x_s+\sum_{k=1}^{n}a_{nk}x_k\end{pmatrix}$$

$$= \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\alpha} + \boldsymbol{A}\boldsymbol{\alpha} = (\boldsymbol{A}^{\mathrm{T}} + \boldsymbol{A})\boldsymbol{\alpha}$$

特别地，当 $\boldsymbol{A}$ 是对称矩阵时，有

$$\frac{\mathrm{d}f}{\mathrm{d}\boldsymbol{\alpha}} = 2\boldsymbol{A}\boldsymbol{\alpha}$$

**定义 6.3.3**　设矩阵 $F(\boldsymbol{A}) = (f_{ij}(\boldsymbol{A}))_{s\times t}$ 的元素 $f_{ij}(\boldsymbol{A})(i = 1, 2, \cdots, s;\ j = 1, 2, \cdots, t)$ 都是矩阵变量 $\boldsymbol{A} = (x_{ij})_{m\times n}$ 的函数，则称 $F(\boldsymbol{A})$ 为矩阵值函数. 规定 $F(\boldsymbol{A})$ 对矩阵变量 $\boldsymbol{A}$ 的导数为

$$\frac{\mathrm{d}F}{\mathrm{d}\boldsymbol{A}} = \begin{pmatrix} \dfrac{\partial F}{\partial x_{11}} & \dfrac{\partial F}{\partial x_{12}} & \cdots & \dfrac{\partial F}{\partial x_{1n}} \\ \dfrac{\partial F}{\partial x_{21}} & \dfrac{\partial F}{\partial x_{22}} & \cdots & \dfrac{\partial F}{\partial x_{2n}} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial F}{\partial x_{m1}} & \dfrac{\partial F}{\partial x_{m2}} & \cdots & \dfrac{\partial F}{\partial x_{mn}} \end{pmatrix} \tag{6.3.4}$$

其中

$$\frac{\partial F}{\partial x_{ij}} = \begin{pmatrix} \dfrac{\partial f_{11}}{\partial x_{ij}} & \dfrac{\partial f_{12}}{\partial x_{ij}} & \cdots & \dfrac{\partial f_{1t}}{\partial x_{ij}} \\ \dfrac{\partial f_{21}}{\partial x_{ij}} & \dfrac{\partial f_{22}}{\partial x_{ij}} & \cdots & \dfrac{\partial f_{2t}}{\partial x_{ij}} \\ \vdots & \vdots & & \vdots \\ \dfrac{\partial f_{s1}}{\partial x_{ij}} & \dfrac{\partial f_{s2}}{\partial x_{ij}} & \cdots & \dfrac{\partial f_{st}}{\partial x_{ij}} \end{pmatrix}$$

即其结果为 $(ms)\times(nt)$ 矩阵.

又由于向量是特殊的矩阵，所以向量值函数对于向量变量的导数，向量值函数对于矩阵变量的导数，矩阵值函数对于向量变量的导数都可以视为定义 6.3.3 的特殊情形来计算.

**例 6.3.8**　设 $\boldsymbol{\alpha} = (x_1, x_2, \cdots, x_n)^{\mathrm{T}}$，求向量值函数 $\boldsymbol{\alpha}^{\mathrm{T}} = (x_1, x_2, \cdots, x_n)$ 对向量变量 $\boldsymbol{\alpha}$ 的导数 $\dfrac{\mathrm{d}\boldsymbol{\alpha}^{\mathrm{T}}}{\mathrm{d}\boldsymbol{\alpha}}$.

**解**

$$\frac{\mathrm{d}\boldsymbol{\alpha}^{\mathrm{T}}}{\mathrm{d}\boldsymbol{\alpha}}=\begin{pmatrix}\frac{\partial\boldsymbol{\alpha}^{\mathrm{T}}}{\partial x_1}\\\frac{\partial\boldsymbol{\alpha}^{\mathrm{T}}}{\partial x_2}\\\vdots\\\frac{\partial\boldsymbol{\alpha}^{\mathrm{T}}}{\partial x_n}\end{pmatrix}=\begin{pmatrix}1&&&\\&1&&\\&&\ddots&\\&&&1\end{pmatrix}=\boldsymbol{E}$$

**例 6.3.9** 设 $\boldsymbol{\alpha}=(a_1,a_2,a_3,a_n)^{\mathrm{T}}$ 是给定向量，$\boldsymbol{A}=(x_{ij})_{2\times4}$ 是矩阵变量，试求 $\frac{\mathrm{d}(\boldsymbol{A\alpha})^{\mathrm{T}}}{\mathrm{d}\boldsymbol{A}}$，$\frac{\mathrm{d}(\boldsymbol{A\alpha})}{\mathrm{d}\boldsymbol{A}}$.

**解** 因为

$$\boldsymbol{A\alpha}=\begin{pmatrix}\sum_{k=1}^{4}x_{1k}a_k\\\sum_{k=1}^{4}x_{2k}a_k\end{pmatrix},\qquad(\boldsymbol{A\alpha})^{\mathrm{T}}=\left(\sum_{k=1}^{4}x_{1k}a_k,\ \sum_{k=1}^{4}x_{2k}a_k\right)$$

所以

$$\frac{\mathrm{d}(\boldsymbol{A\alpha})^{\mathrm{T}}}{\mathrm{d}\boldsymbol{A}}=\begin{pmatrix}\frac{\partial(\boldsymbol{A\alpha})^{\mathrm{T}}}{\partial x_{11}}&\frac{\partial(\boldsymbol{A\alpha})^{\mathrm{T}}}{\partial x_{12}}&\frac{\partial(\boldsymbol{A\alpha})^{\mathrm{T}}}{\partial x_{13}}&\frac{\partial(\boldsymbol{A\alpha})^{\mathrm{T}}}{\partial x_{14}}\\\frac{\partial(\boldsymbol{A\alpha})^{\mathrm{T}}}{\partial x_{21}}&\frac{\partial(\boldsymbol{A\alpha})^{\mathrm{T}}}{\partial x_{22}}&\frac{\partial(\boldsymbol{A\alpha})^{\mathrm{T}}}{\partial x_{23}}&\frac{\partial(\boldsymbol{A\alpha})^{\mathrm{T}}}{\partial x_{24}}\end{pmatrix}$$

$$=\begin{pmatrix}a_1&0&a_2&0&a_3&0&a_4&0\\0&a_1&0&a_2&0&a_3&0&a_4\end{pmatrix}$$

且

$$\frac{\mathrm{d}(\boldsymbol{A\alpha})}{\mathrm{d}\boldsymbol{A}}=\begin{pmatrix}\frac{\partial(\boldsymbol{A\alpha})}{\partial x_{11}}&\frac{\partial(\boldsymbol{A\alpha})}{\partial x_{12}}&\frac{\partial(\boldsymbol{A\alpha})}{\partial x_{13}}&\frac{\partial(\boldsymbol{A\alpha})}{\partial x_{14}}\\\frac{\partial(\boldsymbol{A\alpha})}{\partial x_{21}}&\frac{\partial(\boldsymbol{A\alpha})}{\partial x_{22}}&\frac{\partial(\boldsymbol{A\alpha})}{\partial x_{23}}&\frac{\partial(\boldsymbol{A\alpha})}{\partial x_{24}}\end{pmatrix}=\begin{pmatrix}a_1&a_2&a_3&a_4\\0&0&0&0\\0&0&0&0\\a_1&a_2&a_3&a_4\end{pmatrix}$$



## 6.4 矩阵函数的应用

本节介绍矩阵函数及微积分的一些应用.

在线性控制系统中，常常涉及求解线性微分方程组的问题. 矩阵函数在其中有着重要作用.

**定理 6.4.1** 设 $\boldsymbol{A}$ 是 $n$ 阶常系数矩阵，则满足初始条件的线性齐次微分方程组

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t)\\ \boldsymbol{x}(t_0)=\boldsymbol{x}_0\end{cases}$$

的解为

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0$$

这里

$$\boldsymbol{A}=(a_{ij})_{n\times n},\ \boldsymbol{x}(t)=(x_1(t),\ x_2(t),\ \cdots,\ x_n(t))^{\mathrm{T}}$$

$$\boldsymbol{x}(t_0)=\boldsymbol{x}_0=(x_1^0,\ x_2^0,\ \cdots,\ x_n^0)^{\mathrm{T}}$$

特别地，当 $t_0=0$ 时

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}_0$$

**证** 由于

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t))&=\mathrm{e}^{-\boldsymbol{A}t}(-\boldsymbol{A})\boldsymbol{x}(t)+\mathrm{e}^{-\boldsymbol{A}t}\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t)\\&=\mathrm{e}^{-\boldsymbol{A}t}\left(\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{x}(t)-\boldsymbol{A}\boldsymbol{x}(t)\right)=\boldsymbol{0}\end{aligned}$$

将上式两端在$[t_0,\ t]$上积分，得

$$\int_{t_0}^{t}\frac{\mathrm{d}}{\mathrm{d}t}[\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(\tau)]\mathrm{d}\tau=\boldsymbol{0}$$

$$\mathrm{e}^{-\boldsymbol{A}t}\boldsymbol{x}(t)-\mathrm{e}^{\boldsymbol{A}t_0}\boldsymbol{x}(t_0)=\boldsymbol{0}$$

故微分方程组满足初始条件的解的形式为

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}(t-t_0)}\boldsymbol{x}_0$$

□

若 $t_0=0$，则

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}_0$$

**例 6.4.1** 设 $\boldsymbol{A}=\begin{pmatrix}2&0&0\\1&1&1\\1&-1&3\end{pmatrix}$，求微分方程组 $\dfrac{\mathrm{d}x}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t)$ 满足初始条件

$\boldsymbol{x}(0)=(1, 1, 1)^{\mathrm{T}}$ 的解.

**解** $\boldsymbol{A}$ 的最小多项式为 $m_{A}(\lambda)=(\lambda-2)^{2}$，令

$$r(\lambda)=a_{0}+a_{1}\lambda, \qquad f(\lambda)=\mathrm{e}^{\lambda t}$$

则由

$$f(2)=\mathrm{e}^{2t}=a_{0}+2a_{1}=r(2)$$

$$f'(2)=t\mathrm{e}^{2t}=a_{1}=r'(2)$$

得

$$a_{0}=\mathrm{e}^{2t}-2t\mathrm{e}^{2t}, \qquad a_{1}=t\mathrm{e}^{2t}$$

从而

$$\mathrm{e}^{\boldsymbol{A}t}=a_{0}\boldsymbol{E}+a_{1}\boldsymbol{A}=\begin{bmatrix}\mathrm{e}^{2t} & 0 & 0\\ t\mathrm{e}^{2t} & (1-t)\mathrm{e}^{2t} & t\mathrm{e}^{2t}\\ t\mathrm{e}^{2t} & -\mathrm{e}^{2t} & (1+t)\mathrm{e}^{2t}\end{bmatrix}$$

故微分方程组的解为

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0)=(\mathrm{e}^{2t}, (1+t)\mathrm{e}^{2t}, (1+t)\mathrm{e}^{2t})^{\mathrm{T}}$$

下面我们讨论一阶线性常系数非齐次微分方程组的定解问题.

**定理 6.4.2** 设 $\boldsymbol{A}$ 为 $n$ 阶常系数矩阵，则满足初始条件的线性非齐次微分方程组

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{f}(t)\\ \boldsymbol{x}(t_{0})=x_{0}\end{cases}$$

的解为

$$\boldsymbol{x}(t)=\mathrm{e}^{\boldsymbol{A}(t-t_{0})}\boldsymbol{x}_{0}+\int_{t_{0}}^{t}\mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{f}(\tau)\mathrm{d}\tau \qquad (6.4.1)$$

这里

$$\boldsymbol{f}(t)=(f_{1}(t), f_{2}(t), \cdots, f_{n}(t))^{\mathrm{T}}$$

**证** 改写方程为

$$\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}-\boldsymbol{A}\boldsymbol{x}=\boldsymbol{f}(t)$$

并以 $\mathrm{e}^{-\boldsymbol{A}t}$ 左乘方程两边，得

$$e^{-At}\left(\frac{dx}{dt}-Ax\right)=e^{-At}f(t)$$

即
$$\frac{d}{dt}(e^{-At}x)=e^{-At}f(t)$$

在$[t_0, t]$上进行积分，得

$$e^{-At}x-e^{-At_0}x(t_0)=\int_{t_0}^{t}e^{-A\tau}f(\tau)d\tau$$

即

$$x(t)=e^{A(t-t_0)}x(t_0)+\int_{t_0}^{t}e^{-A\tau}f(\tau)d\tau$$ □

**例 6.4.2**　设$A=\begin{bmatrix}2 & 0 & 0\\ 1 & 1 & 1\\ 1 & -1 & 3\end{bmatrix}$，$f(t)=(e^{2t}, e^{2t}, 0)^T$，求微分方程组$\frac{dx}{dt}=Ax(t)+f(t)$满足初始条件$x(0)=(-1, 1, 0)^T$的解.

**解**　由例 6.4.1 知

$$e^{At}=\begin{bmatrix}e^{2t} & 0 & 0\\ te^{2t} & (1-t)e^{2t} & te^{2t}\\ te^{2t} & -e^{2t} & (1+t)e^{2t}\end{bmatrix}$$

计算

$$e^{-As}f(s)=(1, 1, 0)^T$$

$$\int_0^t e^{-As}f(s)ds=(t, t, 0)^T$$

$$e^{At}x(0)=(-e^{2t}, (1-2t)e^{2t}, -2te^{2t})^T$$

从而

$$\begin{aligned}x(t)&=e^{At}x(0)+\int_0^t e^{A(t-s)}f(s)ds\\&=e^{At}x(0)+e^{At}\int_0^t e^{A(t-s)}f(s)ds\\&=((t-1)e^{2t}, (1-t)e^{2t}, -2te^{2t})^T\end{aligned}$$

以上介绍了矩阵函数在一阶常系数齐次微分方程组及非齐次微分方程组中的应用，这是比较简单的情形. 同样地，在一阶线性变系数齐次微分方程组及非齐次

微分方程组中，矩阵函数都有应用，但都比上面所讲的困难多了. 在此不作更深入的讨论.

矩阵函数在线性系统理论中有广泛的应用. 这里简单讨论一下现代控制论中的两个基本问题：系统的能控性与能观测性. 我们只就连续型的线性定常系统进行讨论.

**定义 6.4.1** 线性方程组

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\boldsymbol{A}\boldsymbol{x}(t)+\boldsymbol{B}\boldsymbol{u}(t) & (6.4.2)\\ \boldsymbol{y}(t)=\boldsymbol{C}\boldsymbol{x}(t)+\boldsymbol{D}\boldsymbol{u}(t) & (6.4.3)\end{cases}$$

称为定常线性系统的状态空间表达式.

微分方程 6.4.2 称为状态方程，变换方程 6.4.3 称为输出方程. 其中：

$\boldsymbol{A}\in\mathbf{C}^{n\times n}$ 称为系统矩阵，表示系统内部各状态变量之间的关联情况；

$\boldsymbol{B}\in\mathbf{C}^{n\times n}$ 称为输入矩阵，表示输入对每个状态变量的作用情况；

$\boldsymbol{C}\in\mathbf{C}^{p\times n}$ 称为输出矩阵或量测矩阵，表示输出与每个状态变量的组成关系；

$\boldsymbol{D}\in\mathbf{C}^{p\times m}$ 称为直接传递矩阵；

$\boldsymbol{x}(t)\in\mathbf{C}^{n\times 1}$ 称为状态向量；

$\boldsymbol{u}(t)\in\mathbf{C}^{m\times 1}$ 称为输入或控制向量；

$\boldsymbol{y}(t)\in\mathbf{C}^{p\times 1}$ 称为输出向量.

为分析简便，往往不考虑输入对输出的直接传递，故 $\boldsymbol{D}=\boldsymbol{0}$，这个系统简记为 $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C})$.

**定义 6.4.2** 对于一个线性定常系统，若在某个有限时间区间 $[0, t_1]$ 内存在着输入 $\boldsymbol{u}(t)(0\leqslant t\leqslant t_1)$，能使系统从任意初始状态 $\boldsymbol{x}(0)=\boldsymbol{x}_0$ 转移到 $\boldsymbol{x}(t_1)=\boldsymbol{\theta}$，则称该状态 $\boldsymbol{x}_0$ 是能控的；若系统的所有状态都是能控的，则称该系统是完全能控的.

**定理 6.4.3** 系统 $(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C})$ 完全能控的充分必要条件是对称矩阵

$$\boldsymbol{W}_C(0, t_1)=\int_0^{t_1}\mathrm{e}^{-\boldsymbol{A}\tau}\boldsymbol{B}\boldsymbol{B}^{\mathrm{T}}\mathrm{e}^{-\boldsymbol{A}^{\mathrm{T}}\tau}\mathrm{d}\tau$$

为非奇异矩阵.

**证** 先证充分性.

设 $\boldsymbol{W}_C(0, t_1)$ 非奇异. 对方程(6.4.2)从 0 到 $t_1$ 积分得

$$\boldsymbol{x}(t_1)=\mathrm{e}^{\boldsymbol{A}t_1}\boldsymbol{x}(0)+\int_0^{t_1}\mathrm{e}^{\boldsymbol{A}(t_1-\tau)}\boldsymbol{B}\boldsymbol{u}(\tau)\mathrm{d}\tau \tag{6.4.4}$$

令

$$u(t) = -B^{T}e^{-A_{t}^{T}}W_{C}^{-1}(0,\ t_1)x(0) \tag{6.4.5}$$

将 $u(t)$ 代入式 6.4.4 得

$$\begin{aligned}
x(t_1) &= e^{At_1}x(0) - e^{At_1}\left(\int_0^{t_1} e^{-A\tau}BB^{T}e^{-A_{\tau}^{T}}d\tau\right)W_{C}^{-1}(0,\ t_1)x(0) \\
&= e^{At_1}x(0) - e^{At_1}W_{C}(0,\ t_1)W_{C}^{-1}(0,\ t_1)x(0) \\
&= e^{At_1}x(0) - e^{At_1}x(0) = \mathbf{0}
\end{aligned}$$

这说明在式(6.4.5)所示的控制输入 $u(t)$ 作用下，能使系统从 $x(0)$ 转移到 $x(t_1) = \mathbf{0}$.

再证必要性. 用反证法.

若系统是完全能控的，但 $W_C(0,\ t_1)$ 是奇异的，则引出矛盾.

因 $W_C(0,\ t_1)$ 是奇异的，则必有非零向量 $\alpha = (a_1,\ a_2,\ \cdots,\ a_n)^{T}$，使对任意时刻 $t = t_1 \geqslant 0$，下式能够成立

$$\alpha^{T}W_{C}(0,\ t_1)\alpha = \mathbf{0}$$

即

$$\int_0^{t_1} \alpha^{T}(e^{-A\tau}BB^{T}e^{-A^{T}\tau})\alpha d\tau = 0$$

$$\int_0^{t_1} (\alpha^{T}e^{-A\tau}B)(\alpha^{T}e^{-A\tau}B)^{T}d\tau = 0$$

故对任意时刻 $t$，有

$$\alpha^{T}e^{-At}B = \mathbf{0},\quad t \geqslant 0$$

又系统是完全能控的，故由充分性，必存在某个 $u(t)$，使其作用于系统上，使 $x(t_1) = \mathbf{0}$. 故由式(6.4.4)得

$$-x(0) = \int_0^{t_1} e^{-A\tau}Bu(\tau)d\tau$$

上式两边左乘以 $\alpha^{T}$，便有

$$-\alpha^{T}x(0) = \int_0^{t_1} \alpha^{T}e^{-A\tau}Bu(\tau)d\tau = \mathbf{0}$$

由 $x(0)$ 为任意的，现选初始状态 $x(0) = \alpha$，则有 $\alpha^{T}\alpha = \mathbf{0}$. 这与 $\alpha$ 是非零向量矛盾，故 $W_C(0,\ t_1)$ 是非奇异的. 定理得到证明.

定理 6.4.3 从理论上给出系统是否完全能控的判别准则，但在实际应用中有些难度. 下面提供一个较简便的判别准则.

**定理 6.4.4** 系统$(A, B, C)$是完全能控的充要条件是 $n\times mn$ 矩阵

$$W_C=(B, AB, A^2B, \cdots, A^{n-1}B)$$

的秩为 $n$(行满秩矩阵). $W_C$ 称为能控性矩阵.

**证** 这个定理可以用反证法证明. 这里只给出充分性的证明. 如果系统不是完全能控的,则由定理 6.4.3 知矩阵 $W_C(0, t_1)$是奇异的,故存在非零 $\alpha$, 使得

$$\alpha^{T}e^{-At}B=0$$

两边求 $k$ 次微商可得

$$\alpha^{T}(-A)^{k}e^{-At}B=0,\quad k=1, 2, \cdots, n-1$$

令 $t=0$ 得

$$\alpha^{T}A^{k}B=0,\quad k=1, 2, \cdots, n-1$$

而当 $t=0$ 时,由

$$\alpha^{T}e^{-At}B=0$$

得

$$\alpha^{T}B=0$$

故对 $k=0, 1, 2, \cdots, n-1$, 都有

$$\alpha^{T}A^{k}B=0$$

即

$$\alpha^{T}(B, AB, A^2B, \cdots, A^{n-1}B)=0\ (\alpha\text{ 是非零向量})$$

这与矩阵 $W_C$ 的秩为 $n$ 相矛盾. 这就证明了系统是完全能控的.

**例 6.4.3** 设 $A=\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$, $B=\begin{pmatrix}1\\ 0\end{pmatrix}$, 判断系统$(A, B, C)$是否完全能控.

**解**
$$W_C=(B, AB)=\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix}$$

这时 $W_C$ 的秩 $=1\neq 2$, 故系统$(A, B, C)$不是完全能控的.

**例 6.4.4** 设

$$A=\begin{bmatrix}0 & 0 & -3\\ 2 & 0 & -7\\ 0 & -1 & 0\end{bmatrix},\qquad B=\begin{bmatrix}1 & 1\\ 0 & -1\\ 0 & 1\end{bmatrix}$$

判断系统$(\boldsymbol{A},\ \boldsymbol{B},\ \boldsymbol{C})$是否完全能控.

**解**

$$\boldsymbol{W}_C=(\boldsymbol{B},\ \boldsymbol{AB},\ \boldsymbol{A}^2\boldsymbol{B})=\begin{pmatrix}1 & 1 & 0 & -3 & 0 & -3\\ 0 & -1 & 2 & -5 & 0 & -13\\ 0 & 1 & 0 & 1 & -2 & 5\end{pmatrix}$$

这时$\boldsymbol{W}_C$的秩$=3$,故系统$(\boldsymbol{A},\ \boldsymbol{B},\ \boldsymbol{C})$是完全能控的.

下面介绍能观测性问题.

**定义 6.4.3**　对于一个线性定常系统,若在有限时间区间$[0,\ t_1]$内,能通过观测系统的输出$\boldsymbol{y}(t)$而唯一地确定任意初始状态$\boldsymbol{x}(0)$,则称该系统是完全能观测的,或称对每一状态$\boldsymbol{x}(0)$是能观测的.

**定理 6.4.5**　系统$(\boldsymbol{A},\ \boldsymbol{B},\ \boldsymbol{C})$完全能观测的充要条件是$n$阶对称矩阵

$$\boldsymbol{M}(0,\ t_1)=\int_0^{t_1}\mathrm{e}^{\boldsymbol{A}^{\mathrm{T}}\tau}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}\tau}\mathrm{d}\tau \tag{6.4.6}$$

是非奇异矩阵.

**证**　先证充分性.

$$\begin{aligned}\boldsymbol{y}(t)&=\boldsymbol{Cx}(t)+\boldsymbol{Du}(t)\\ &=\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0)+\boldsymbol{C}\int_0^t\mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{Bu}(\tau)\mathrm{d}\tau+\boldsymbol{Du}(t)\end{aligned}$$

令

$$\boldsymbol{\eta}(t)=\boldsymbol{y}(t)-\boldsymbol{C}\int_0^t\mathrm{e}^{\boldsymbol{A}(t-\tau)}\boldsymbol{Bu}(\tau)\mathrm{d}\tau-\boldsymbol{Du}(t)$$

于是得

$$\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0)=\boldsymbol{\eta}(t) \tag{6.4.7}$$

以$\mathrm{e}^{\boldsymbol{A}^{\mathrm{T}}t}\boldsymbol{C}^{\mathrm{T}}$左乘上式两边,并从 0 到$t_1$进行积分,得

$$\left(\int_0^{t_1}\mathrm{e}^{\boldsymbol{A}^{\mathrm{T}}t}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}t}\mathrm{d}t\right)\boldsymbol{x}(0)=\int_0^{t_1}\mathrm{e}^{\boldsymbol{A}^{\mathrm{T}}t}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\eta}(t)\mathrm{d}t$$

即

$$\boldsymbol{M}(0,\ t_1)\boldsymbol{x}(0)=\int_0^{t_1}\mathrm{e}^{\boldsymbol{A}^{\mathrm{T}}t}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\eta}(t)\mathrm{d}t$$

由于$\boldsymbol{M}(0,\ t_1)$是非奇异矩阵,则

$$\boldsymbol{x}(0)=\boldsymbol{M}^{-1}(0,\ t_1)\int_0^{t_1}\mathrm{e}^{\boldsymbol{A}^{\mathrm{T}}t}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\eta}(t)\mathrm{d}t$$

由此,$\boldsymbol{x}(0)$可以唯一确定.故系统是完全能观测的.

再证必要性. 用反证法. 设系统是完全能观测的,若 $\boldsymbol{M}(0,\ t_1)$是奇异的,则存在非零 $n$ 维向量 $\boldsymbol{\alpha}$,使得对任意的 $t\geqslant 0$, 有

$$\boldsymbol{\alpha}^{\mathrm{T}}\boldsymbol{M}(0,\ t_1)\boldsymbol{\alpha}=\boldsymbol{0}$$

即

$$\boldsymbol{\alpha}^{\mathrm{T}}\left(\int_0^{t_1}\mathrm{e}^{\boldsymbol{A}^{\mathrm{T}}\tau}\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}\tau}\,\mathrm{d}\tau\right)\boldsymbol{\alpha}=\boldsymbol{0}$$

$$\int_0^{t_1}(\boldsymbol{\alpha}^{\mathrm{T}}\mathrm{e}^{\boldsymbol{A}^{\mathrm{T}}\tau}\boldsymbol{C}^{\mathrm{T}})(\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}\tau}\boldsymbol{\alpha})\,\mathrm{d}\tau=\boldsymbol{0}$$

这表明对任意的 $t\geqslant 0$, 有

$$\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{\alpha}=\boldsymbol{0}$$

取 $\boldsymbol{\alpha}=\boldsymbol{x}(0)\neq\boldsymbol{0}$,则 $\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0)=\boldsymbol{0}$. 这与当 $\boldsymbol{x}(0)=\boldsymbol{0}$ 时,由式(6.4.7)

$$\boldsymbol{\eta}(t)=\boldsymbol{C}\mathrm{e}^{\boldsymbol{A}t}\boldsymbol{x}(0)=\boldsymbol{0}$$

的结果作比较,说明 $\boldsymbol{x}(0)$不能唯一确定. 这与系统是完全能观测矛盾,故 $\boldsymbol{M}(0,\ t_1)$是非奇异矩阵. □

**定理 6.4.6** 系统($\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$)完全能观测的充要条件是 $pn\times n$ 矩阵

$$\boldsymbol{W}_0=\begin{bmatrix}\boldsymbol{C}\\ \boldsymbol{C}\boldsymbol{A}\\ \vdots\\ \boldsymbol{C}\boldsymbol{A}^{n-1}\end{bmatrix}$$

的秩等于 $n$(列满秩矩阵). 矩阵 $\boldsymbol{W}_0$ 称为能观测性矩阵.

**例 6.4.5** 判定线性系统

$$\begin{cases}\dfrac{\mathrm{d}\boldsymbol{x}(t)}{\mathrm{d}t}=\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}\boldsymbol{x}(t)+\begin{pmatrix}2 & -1\\ 0 & 1\end{pmatrix}\boldsymbol{u}(t)\\ \boldsymbol{y}(t)=\begin{pmatrix}1 & 1\\ 2 & 3\end{pmatrix}\boldsymbol{x}(t)\end{cases}$$

的能观测性.

**解** 能观性矩阵为

$$\boldsymbol{W}_0=\begin{pmatrix}\boldsymbol{C}\\ \boldsymbol{C}\boldsymbol{A}\end{pmatrix}=\begin{bmatrix}1 & -1\\ 2 & -2\\ 0 & 2\\ 0 & 4\end{bmatrix}$$

所以 $\boldsymbol{W}_0$ 的秩为 2，故系统是能观测的.

**定义 6.4.4**　设有两个线性系统$(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C})$和$(\tilde{\boldsymbol{A}},\tilde{\boldsymbol{B}},\tilde{\boldsymbol{C}})$，如果存在 $n$ 阶可逆矩阵 $\boldsymbol{P}$，使得

$$\tilde{\boldsymbol{A}}=\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$$

$$\tilde{\boldsymbol{B}}=\boldsymbol{P}^{-1}\boldsymbol{B}$$

$$\tilde{\boldsymbol{C}}=\boldsymbol{C}\boldsymbol{P}$$

则称系统$(\tilde{\boldsymbol{A}},\tilde{\boldsymbol{B}},\tilde{\boldsymbol{C}})$与系统$(\boldsymbol{A},\boldsymbol{B},\boldsymbol{C})$代数等价.

显然，代数等价具有自反性、对称性和传递性.

易证下列定理成立.

**定理 6.4.7**　两个代数等价系统具有相同的能控性和能观测性.

## 习 题 6

1. 判断下列矩阵是否为收敛矩阵：

(1) $\boldsymbol{A}=\dfrac{1}{6}\begin{pmatrix}1 & -8\\ -2 & 1\end{pmatrix}$；

(2) $\boldsymbol{A}=\begin{pmatrix}0.1 & 0.2 & -0.2\\ -0.4 & 0.3 & 0.1\\ 0.3 & 0.1 & 0.2\end{pmatrix}$.

2. 已知

(1) $\boldsymbol{A}^{(k)}=\begin{pmatrix}\dfrac{1}{2^k} & \dfrac{1}{3^k}\\ \dfrac{1}{(k+1)(k+2)} & 0\end{pmatrix}$；　(2) $\boldsymbol{A}^{(k)}=\begin{pmatrix}0.1 & 0.7\\ 0.3 & 0.6\end{pmatrix}^k$.

试判断矩阵级数 $\sum\limits_{k=0}^{+\infty}\boldsymbol{A}^{(k)}$ 的敛散性.

3. 设

$$\boldsymbol{A}=\begin{pmatrix}2 & 0 & 0\\ 1 & 1 & 1\\ 1 & -1 & 3\end{pmatrix}$$

试利用 Jordan 标准形方法计算 $e^{\boldsymbol{A}}$，$e^{\boldsymbol{A}t}$，$\sin\boldsymbol{A}$.

4. 设

$$A = \begin{pmatrix} 2 & 1 & 4 \\ 0 & 2 & 0 \\ 0 & 3 & 1 \end{pmatrix}$$

试利用待定系数法计算 $e^{At}$，$\sin At$.

5. 设

$$A = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}$$

试求 $\ln A$.

6. 设

$$A(t) = \begin{pmatrix} e^{2t} & te^{t} & 1 \\ e^{-t} & 2e^{2t} & 0 \\ 3t & 0 & 0 \end{pmatrix}$$

试求：(1) $\frac{d}{dt}A(t)$，$\frac{d}{dt}\mid A(t)\mid$；(2) $\int A(t)dt$；$\int_0^1 A(t)dt$.

7. 证明：

(1) 若 $A$ 为实反对称阵，则 $e^{A}$ 为正交阵；

(2) 若 $A$ 为厄米特矩阵，则 $e^{iA}$ 为酉矩阵.

8. 设 $A=(x_{ij})_{n\times n}$，试求 $\frac{d}{dA}|A|$.

9. 设 $A$ 是 $n$ 阶实对称阵，$x$，$b\in R^n$，$f(x)=\frac{1}{2}x^TAx-b^Tx$，求 $\frac{df}{dx}$.

10. 设 $f(A)=\mathrm{tr}(A^TA)$，其中 $A\in R^{m\times n}$ 是矩阵变量，试求 $\frac{df}{dA}$.

11. 设 $A\in C^{m\times n}$，$\alpha\in R^n$ 是向量变量，$F(\alpha)=A\alpha$，试求 $\frac{dF(\alpha)}{d\alpha}$.

12. 设 $A$ 是 $n\times m$ 常数矩阵，$\alpha=(x_1, x_2, \cdots, x_n)^T$ 是向量变量，且 $F(\alpha)=\alpha^TA$，试求 $\frac{dF(\alpha)}{d\alpha}$.

13. 试求微分方程组 $\frac{dx(t)}{dt}=Ax(t)$ 满足初始条件 $x(0)=x_0$ 的解，其中

$$A=\begin{pmatrix}1 & 2 & 1\\ 1 & -1 & 1\\ 2 & 0 & 1\end{pmatrix},\qquad x(0)=(1,\ 0,\ 0)^{T}$$

14. 试求解微分方程组初值问题

$$\begin{cases}\dfrac{dx_1}{dt}=-x_1+x_3+1\\ \dfrac{dx_2}{dt}=x_1+2x_2-1\\ \dfrac{dx_3}{dt}=-4x_1+3x_3+2\\ x_1(0)=1,\ x_2(0)=0,\ x_3(0)=1\end{cases}$$

15. 判断线性系统($\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{C}$)的能控性：

$$\frac{d\boldsymbol{x}(t)}{dt}=\begin{pmatrix}1 & 1 & 0\\ 0 & 1 & 0\\ 0 & 1 & 1\end{pmatrix}\boldsymbol{x}(t)+\begin{pmatrix}0 & 1\\ 1 & 0\\ 0 & 1\end{pmatrix}\boldsymbol{u}(t)$$

16. 判断线性系统($\boldsymbol{A}$，$\boldsymbol{B}$，$\boldsymbol{C}$)的能观测性：

$$\begin{cases}\dfrac{d\boldsymbol{x}(t)}{dt}=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}\boldsymbol{x}(t)+\begin{pmatrix}1\\ -1\end{pmatrix}\boldsymbol{u}(t)\\ \boldsymbol{y}(t)=\begin{pmatrix}0 & 0\\ 2 & 0\end{pmatrix}\boldsymbol{x}(t)\end{cases}$$

# 第7章 矩阵特征值的界 非负矩阵

如果要求出一个矩阵的精确的特征值，通常是一件困难的工作，因为当矩阵的阶数较高时，必须求解高次代数方程. 然而，在矩阵的理论研究及工程计算中，我们常常只需知道特征值在什么范围内变化或确定特征值在复平面上的变化区域. 比如在数值计算中，如果用迭代法讨论线性方程组解的敛散性时，要估计系数矩阵的特征值是否在以原点为圆心的单位圆内；在判断方阵的幂级数是否收敛也要看方阵的特征值的模是否小于某一正实数. 在扰动理论中常要考虑特征值的变化范围. 确定特征值的界，就是要找到复平面上的区域，使得矩阵的特征值在该区域内. 区域的直径越小，特征值的界定范围就越小. 本章将介绍圈定这类区域的 Geršgorin 圆盘定理、Courant-Fischer（柯朗-费歇尔）定理等.

本章要讨论的另一类问题是非负矩阵. 非负矩阵有着实际的应用背景，在许多科学技术领域都有不同程度的应用. 本章介绍正矩阵、非负矩阵、随机矩阵和 M 矩阵的一些重要性质.

## 7.1 Geršgorin 定理

有各种计算方法确定矩阵特征值的变化范围，其中最容易计算、也最实用的一个结果是 1931 年由 Geršgorin 所给出的盖尔圆盘.

**定义 7.1.1** 设 $\boldsymbol{A}=(a_{ij})\in\mathbf{C}^{n\times n}$，令 $R_i=\sum\limits_{\substack{j=1\\j\neq i}}^{n}|a_{ij}|$，$i=1,2,\cdots,n$. 称复平面上的闭圆盘

$$\{z\mid |z-a_{ii}|\leqslant R_i\},\quad i=1,2,\cdots,n$$

为矩阵 $\boldsymbol{A}$ 的行盖尔圆. 类似地可以定义矩阵 $\boldsymbol{A}$ 的 $n$ 个列盖尔圆.

在给出主要结果之前，我们指出要用到的一个代数学上的结果，即矩阵的特征值是其元素的连续函数. 这是因为，特征值是特征多项式的零点，因而是特征多项式系数的连续函数. 但特征多项式的系数（由计算公式可知）又是 $\boldsymbol{A}$ 的元素的连续

函数. 所以,当 $\boldsymbol{A}$ 的系数连续变化时,必有对应特征值的连续变化.

**定理 7.1.1**　设 $\boldsymbol{A} \in \mathbf{C}^{n\times n}$, $D_1$, $D_2$, $\cdots$, $D_n$是$\boldsymbol{A}$ 的 $n$ 个行盖尔圆,$\lambda$ 为$\boldsymbol{A}$ 的任一特征值,则

$$\lambda \in \bigcup_{i=1}^{n} D_i \tag{7.1.1}$$

若由 $m$ 个盖尔圆构成的并集$G$ 与其余的$n-m$个盖尔圆不交,则 $G$ 中恰有$\boldsymbol{A}$ 的$m$ 个特征值.

**证**　设$\lambda$ 是矩阵$\boldsymbol{A}$ 的一个特征值,$\boldsymbol{x}$ 是属于特征值$\lambda$ 的特征向量,则$\boldsymbol{Ax}=\lambda\boldsymbol{x}$. 具体地写出这两个相等向量的分量就是

$$\sum_{k=1}^{n} a_{jk}x_k = \lambda x_j, \quad j=1, 2, \cdots, n \tag{7.1.2}$$

设 $|x_p| = \max\limits_j |x_j|$, 则由以上 $n$ 个等式的第 $p$ 个方程为

$$|\lambda - a_{pp}||x_p| = \left|\sum_{\substack{k=1\\k\neq p}}^{n} a_{pk}x_k\right| \leqslant \sum_{\substack{k=1\\k\neq p}}^{n} |a_{pk}||x_k|$$

$$\leqslant |x_p| \sum_{\substack{k=1\\k\neq p}}^{n} |a_{pk}| = |x_p| R_p$$

因为 $\boldsymbol{x}\neq\boldsymbol{0}$,故 $|x_p|\neq 0$, 从而

$$|\lambda - a_{pp}| \leqslant R_p$$

这就完成了第一个结论的证明.

假定 $m=1$, 即有一个盖尔圆与其他盖尔圆均不交,我们证明在该盖尔圆中有且仅有 $\boldsymbol{A}$ 的一个特征值.

将矩阵 $\boldsymbol{A}$ 写成$\boldsymbol{A}=\boldsymbol{D}+\boldsymbol{C}$,其中$\boldsymbol{D}=\mathrm{diag}(a_{11}, a_{22}, \cdots, a_{nn})$,并设$\boldsymbol{B}(t)=\boldsymbol{D}t+\boldsymbol{C}$,则$\boldsymbol{B}(0)=\boldsymbol{C}$, $\boldsymbol{B}(1)=\boldsymbol{A}$. 当$t$ 在区间$[0, 1]$内连续变化时,必然导致矩阵$\boldsymbol{B}(t)$的特征值的连续变化. 也就是说,对每个 $t\in[0, 1]$, $\boldsymbol{B}(t)$的特征值必在以 $a_{jj}$ 为圆心,以 $tR_j$ $(j=1, 2, \cdots, n)$ 为半径的圆盘内.

不妨设 $\boldsymbol{A}=\boldsymbol{B}(1)$ 的第一个盖尔圆 $D_1$ 与其余的 $n-1$ 个盖尔圆不交,显然,$\boldsymbol{B}(t)$ 的第一个盖尔圆不会与 $\boldsymbol{B}(t)$的其余的盖尔圆相交, $t\in[0, 1]$. 当 $t=0$ 时,$\boldsymbol{B}(0)$ 的特征值是 $a_{11}$, $a_{22}$, $\cdots$, $a_{nn}$. 此时, $a_{11}$是在 $\boldsymbol{B}(0)$的第一个盖尔圆中的唯一一个特征值. 由于$\boldsymbol{B}(t)$的特征值是$t$ 的连续函数,并且$\boldsymbol{B}(t)$的第一个盖尔圆不与其他盖尔圆相交,所以,对任意的 $t\in[0, 1]$, 在 $\boldsymbol{B}(t)$的第一个盖尔圆中有且仅有一个特征值. 特别地,当 $t=1$ 时,在 $\boldsymbol{A}=\boldsymbol{B}(1)$ 的第一个盖尔圆 $D_1$ 中恰有一个特征值.

当 $m \leqslant n$ 时,容易类似地证明定理 7.1.1 的第二个结论. □

将定理 7.1.1 的结论应用到 $\boldsymbol{A}^{\mathrm{T}}$,可以得到关于 $\boldsymbol{A}$ 的列盖尔圆的类似结果.

**例 7.1.1** 估计下列矩阵

$$\boldsymbol{A}=\begin{pmatrix} 0 & 1 & 0 & \mathrm{i} \\ 1 & 6 & 1 & 1 \\ \frac{1}{2}\mathrm{i} & \mathrm{i} & 5\mathrm{i} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} & -2 \end{pmatrix}$$

的特征值分布区域.

**解** 由定理 7.1.1 求出 $\boldsymbol{A}$ 的 4 个行盖尔圆为

$$D_1: |z| \leqslant 2,$$

$$D_2: |z-6| \leqslant 3$$

$$D_3: |z-5\mathrm{i}| \leqslant \frac{3}{2},$$

$$D_4: |z+2| \leqslant 1$$

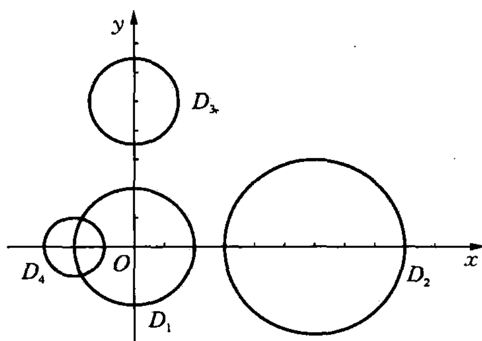


图 7.1.1

则 A 的特征值必在区域 $\bigcup_{i=1}^{4} D_i$ 中,如图 7.1.1 所示.

**例 7.1.2** 设 $\boldsymbol{A}=(a_{ij}) \in \mathbf{C}^{n\times n}$, $R_i$ 是$\boldsymbol{A}$ 的第$i$ 个行盖尔圆半径, $i=1, 2, \cdots, n$. 若对任意的 $i(1 \leqslant i \leqslant n)$, $|a_{ii}|>R_i$, 则称 $\boldsymbol{A}$ 是行对角占优矩阵.

证明:行对角占优矩阵是非奇异的.

**证** 我们只需证明 $\boldsymbol{A}$ 的任一特征值不为零. 设 $\lambda$ 是 $\boldsymbol{A}$ 的任一特征值,则由定理 7.1.1 可知,必有 $1 \leqslant i_0 \leqslant n$, 使得

$$||\lambda|-|a_{i_0 i_0}|| \leqslant |\lambda - a_{i_0 i_0}| \leqslant R_{i_0}$$

从而

$$|\lambda| \geqslant |a_{i_0 i_0}| - R_{i_0} > 0$$

尽管定理 7.1.1 指出当 $m$ 个圆盘的并集 $G$ 与其他盖尔圆不交时,$G$ 中恰有 $m$ 个特征值,但这些特征值具体落在哪几个圆盘中是无法确定的. 比如例 7.1.1 中, $D_1 \cup D_4$ 中有两个特征值,但这两个特征值也许在其中一个圆盘内,也许在 $D_1 \cap D_2$ 内,或 $D_1$, $D_4$ 中各有一个. 各种情况都可能出现. 为解决这一问题,我们

需要用这些圆盘作适当的放缩,使它们彼此隔离.

设矩阵 $\boldsymbol{A}=(a_{ij})_n$ 的 $n$ 个盖尔圆的半径分别为 $R_1, R_2, \cdots, R_n$. 取一组正数 $d_1, d_2, \cdots, d_n$,其中 $d_i \neq 1$, 其余的 $d_j=1, j\neq i$, 令

$$\boldsymbol{D}=\operatorname{diag}(d_1, d_2, \cdots, d_n)$$

则 $\boldsymbol{D}$ 是可逆的,且 $\boldsymbol{B}=\boldsymbol{DAD}^{-1}$ 与 $\boldsymbol{A}$ 有相同的特征值. 若设 $\widetilde{R}_1, \widetilde{R}_2, \cdots, \widetilde{R}_n$ 是矩阵 $\boldsymbol{B}$ 的 $n$ 个盖尔圆的半径,则

$$\widetilde{R}_i=\sum_{\substack{j=1\\ j\neq i}}^{n}\left|a_{ij}\frac{d_i}{d_j}\right|=d_i\sum_{\substack{j=1\\ j\neq i}}^{n}|a_{ij}|\frac{1}{d_j}=d_iR_i \tag{7.1.3}$$

$$\begin{aligned}\widetilde{R}_k&=\sum_{\substack{j=1\\ j\neq k}}^{n}\left|a_{kj}\frac{d_k}{d_j}\right|=d_k\sum_{\substack{j=1\\ j\neq k}}^{n}|a_{kj}|\frac{1}{d_j}\\ &=R_k+\left(\frac{1}{d_i}-1\right)|a_{ki}|\end{aligned} \tag{7.1.4}$$

$(k\neq i;\ k=1, 2, \cdots, n)$, 由式(7.1.3)、式(7.1.4)可以看出,若 $d_i<1$, 则 $\boldsymbol{A}$ 的第 $i$ 个盖尔圆半径缩小,而其余的 $n-1$ 个盖尔圆半径增大;反之,当 $l_i>1$ 时, $\boldsymbol{A}$ 的第 $i$ 个盖尔圆半径增大而其余的圆盘半径缩小. 因此,我们可以通过适当取 $\boldsymbol{D}$ 的方法对 $\boldsymbol{A}$ 的特征值进行隔离.

**例 7.1.3** 对例 7.1.1 中的特征值进行隔离.

**解** 取 $\boldsymbol{D}=\operatorname{diag}\left(\frac{1}{3}, 1, 1, 1\right)$,则矩阵 $\boldsymbol{B}=\boldsymbol{DAD}^{-1}$ 的 4 个盖尔圆圆心不变,而半径分别为

$$\widetilde{R}_1=\frac{1}{3}R_1=\frac{2}{3}, \qquad \widetilde{R}_2=R_2+2|a_{21}|=5$$

$$\widetilde{R}_3=R_3+2|a_{31}|=\frac{5}{2}, \quad \widetilde{R}_4=R_4+2|a_{41}|=1$$

这时,$\boldsymbol{A}$ 中的两个盖尔圆彼此隔离了. 结合例 7.1.1 的结论可知,下列 4 个圆盘中各有 $\boldsymbol{A}$ 的一个特征值

$$|z|\leqslant\frac{2}{3}, \qquad |z-6|\leqslant 3$$

$$|z-5\mathrm{i}|\leqslant\frac{3}{2}, \qquad |z+2|\leqslant 1$$

## 7.2 特征值估计的基本不等式

对任一 $n$ 阶矩阵 $\boldsymbol{A}$，我们总可以将 $\boldsymbol{A}$ 写成 $\boldsymbol{A}=\boldsymbol{B}+\boldsymbol{C}$ 的形式，其中 $\boldsymbol{B}=\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}})$，$\boldsymbol{C}=\frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{H}})$. 如果 $\boldsymbol{C}=\boldsymbol{0}$，即 $\boldsymbol{A}=\boldsymbol{A}^{\mathrm{H}}$，则 $\boldsymbol{A}$ 是 Hermite 矩阵，其特征值必全为实数. 由于矩阵的特征值是其元素的连续函数，可以设想当 $\boldsymbol{C}$ 的元素均在 0 的附近变化时，必然使得 $\boldsymbol{A}=\boldsymbol{B}+\boldsymbol{C}$ 的特征值会出现复数. 所以矩阵 $\boldsymbol{C}$ 应该在确定 $\boldsymbol{A}$ 的特征值的虚部变化范围起着作用. 同理，矩阵 $\boldsymbol{B}$ 应与 $\boldsymbol{A}$ 的特征值的实部变化有某种关联.

**定理 7.2.1** 设 $\boldsymbol{A}\in\mathbf{C}^{n\times n}$，$\lambda_1, \lambda_2, \cdots, \lambda_n$ 是 $\boldsymbol{A}$ 的特征值，则

$$\sum_{i=1}^{n}|\lambda_i|^2\leqslant\|A\|_F^2 \tag{7.2.1}$$

$$\sum_{i=1}^{n}|\mathrm{Re}(\lambda_i)|^2\leqslant\|B\|_F^2 \tag{7.2.2}$$

$$\sum_{i=1}^{n}|\mathrm{Im}(\lambda_i)|^2\leqslant\|C\|_F^2 \tag{7.2.3}$$

其中

$$\boldsymbol{B}=\frac{1}{2}(\boldsymbol{A}+\boldsymbol{A}^{\mathrm{H}}),\qquad \boldsymbol{C}=\frac{1}{2}(\boldsymbol{A}-\boldsymbol{A}^{\mathrm{H}})$$

式(7.2.1)～式(7.2.3)中任一等式成立，必有其余两个等式成立. 而任一等式成立的充分必要条件是 $\boldsymbol{A}$ 为正规矩阵.

**证** 根据 Schur 引理，存在酉矩阵 $\boldsymbol{U}$ 及上三角阵 $\boldsymbol{T}$，使得 $\boldsymbol{A}=\boldsymbol{U}\boldsymbol{T}\boldsymbol{U}^{\mathrm{H}}$. $\boldsymbol{T}$ 的主对角线上的元素是 $\boldsymbol{A}$ 的特征值. 由于矩阵的 $F$-范数在酉变换下不变，所以 $\|\boldsymbol{A}\|_F=\|\boldsymbol{T}\|_F$，记 $\boldsymbol{T}=(t_{rs})$ $(r, s=1, 2, \cdots, n, r>s$ 时，$t_{rs}=0)$，则有

$$\|\boldsymbol{A}\|_F^2=\|\boldsymbol{T}\|_F^2=\sum_{i=1}^{n}|\lambda_i|^2+\sum_{r<s}|t_{rs}|^2\geqslant\sum_{i=1}^{n}|\lambda_i|^2$$

为证式(7.2.2)，注意到 $\boldsymbol{B}=\frac{1}{2}\boldsymbol{U}(\boldsymbol{T}+\boldsymbol{T}^{\mathrm{H}})\boldsymbol{U}^{\mathrm{H}}$，故

$$\|\boldsymbol{B}\|_F^2=\left\|\frac{1}{2}(\boldsymbol{T}+\boldsymbol{T}^{\mathrm{H}})\right\|_F^2=\sum_{i=1}^{n}\left|\frac{\lambda_i+\bar{\lambda}_i}{2}\right|^2+\sum_{r\neq s}\left|\frac{t_{rs}+\bar{t}_{sr}}{2}\right|^2$$

$$= \sum_{i=1}^{n} |\operatorname{Re}(\lambda_i)|^2 + \sum_{r<s} \frac{|t_{rs}|^2}{2} \geqslant \sum_{i=1}^{n} |\operatorname{Re}(\lambda_i)|^2$$

式(7.2.3)可以类似地证明.

从上面的证明过程易知,式(7.2.1)～式(7.2.3)任一等式成立,当且仅当 $t_{rs}=0, r\neq s$; 当且仅当 $\boldsymbol{A}$ 为正规矩阵. □

**推论 7.2.1** 设 $\boldsymbol{A}=(a_{ij}), \boldsymbol{B}=(b_{ij}), \boldsymbol{C}=(c_{ij})$ 为定理 7.2.1 中的矩阵,且

$$\rho = \max_{i,j} |a_{ij}|, \quad \sigma = \max_{i,j} |b_{ij}|, \quad \tau = \max_{i,j} |c_{ij}|$$

则当 $\lambda$ 为 $\boldsymbol{A}$ 的任一特征值时,有

$$|\lambda| \leqslant n\rho, \quad |\operatorname{Re}(\lambda)| \leqslant n\sigma, \quad |\operatorname{Im}(\lambda)| \leqslant n\tau$$

**证** 由定理 7.2.1 知

$$|\lambda|^2 \leqslant \sum_{i=1}^{n} |\lambda_i|^2 \leqslant \|\boldsymbol{A}\|_F^2 = \sum_{i=1}^{n}\sum_{j=1}^{n} |a_{ij}|^2$$
$$\leqslant n^2 \max_{i,j} |a_{ij}| = n^2\rho^2$$

由上式立即可得 $|\lambda| \leqslant n\rho$. 另两个不等式的证明类似. □

由这一推论还可以得到,当 $\boldsymbol{A}$ 是 Hermite 矩阵时,$\boldsymbol{C}=\boldsymbol{0}$,从而 $\tau=0$,此时必有 $\operatorname{Im}(\lambda)=0$,所以 $\boldsymbol{A}$ 的特征值 $\lambda$ 为实数. 同理,当 $\boldsymbol{A}$ 是反 Hermite 矩阵时,$\boldsymbol{B}=\boldsymbol{0}$,此时 $\operatorname{Re}(\lambda)=0$,从而 $\boldsymbol{A}$ 的特征值是 0 或纯虚数.

**推论 7.2.2** 若 $\boldsymbol{A}$ 是 $n$ 阶实矩阵,$\tau = \frac{1}{2}\max_{r,s} |a_{rs} - a_{sr}|$,$\lambda$ 是 $\boldsymbol{A}$ 的任一特征值,则

$$|\operatorname{Im}(\lambda)| \leqslant \tau\sqrt{\frac{n(n-1)}{2}} \tag{7.2.4}$$

**证** 由式(7.2.3)得

$$\sum_{i=1}^{n} |\operatorname{Im}(\lambda_i)|^2 \leqslant \|\boldsymbol{C}\|_F^2 = \sum_{\substack{r,s=1\\ r\neq s}}^{n} \left|\frac{a_{rs}-a_{sr}}{2}\right|^2 \leqslant n(n-1)\tau^2$$

又因 $\boldsymbol{A}$ 是实矩阵,其复特征值成共轭对出现. 所以,当 $\lambda$ 是实特征值时,式(7.2.4)是自然成立的. 而当 $\lambda$ 是复特征值时,有

$$2|\operatorname{Im}(\lambda)| \leqslant \tau^2 n(n-1)$$

这就证明了式(7.2.4)成立. □

一般而言,在估计矩阵的特征值变化范围时,Geršgorin 圆盘定理所得的结果

要比推论 7.2.1 的结论好. 比如用推论 7.2.1 估计例 7.1.1 特征值的变化范围可知

$$|\lambda| \leqslant 24, \quad |\mathrm{Re}(\lambda)| \leqslant 24, \quad |\mathrm{Im}(\lambda)| \leqslant 1$$

这一结果告诉我们,矩阵的特征值在复平面上的闭圆盘与矩形闭区域的交集之中.

$$\{z \mid |z| \leqslant 24\} \cap \{z \mid |\mathrm{Re}(z)| \leqslant 24, |\mathrm{Im}(z)| \leqslant 1\}$$

与例 7.2.1 的估计区域比较优劣立判.

不过,推论 7.2.2 的结论对某些特殊的矩阵,如实反对称阵还是较理想的.

**例 7.2.1** 设三阶实反对称阵

$$\boldsymbol{A} = \begin{pmatrix} 0 & 0.2 & 0.1 \\ -0.2 & 0 & 0.2 \\ -0.1 & -0.2 & 0 \end{pmatrix}$$

试估计 $\boldsymbol{A}$ 的特征值的分布区域.

**解** 由于 $\boldsymbol{A}$ 是实反对称阵,故其特征值是 0 或纯虚数. 由推论 7.2.2 计算得知

$$|\mathrm{Im}(\lambda)| \leqslant \tau\sqrt{\frac{n(n-1)}{2}} = 0.2\sqrt{3} \approx 0.3464$$

即 $\boldsymbol{A}$ 的任一特征值 $\lambda = a\mathrm{i}$ 满足关系式 $|a| \leqslant 0.3464$. 实际求出 $\boldsymbol{A}$ 的特征值为

$$\lambda_1 = 0, \quad \lambda_2 = -0.3\mathrm{i}, \quad \lambda_3 = 0.3\mathrm{i}$$

## 7.3 Courant-Fischer 定理和 Hermite 矩阵的特征值

在这一节,我们来讨论 Hermite 矩阵的特征值的一种计算方法. 这一方法是在 1905 年由 E. Fischer 发现而在 1920 年由 R. Courant 整理完善的. 我们约定,本节中关于复向量的内积均是标准内积.

**定义 7.3.1** 设 $\boldsymbol{H}$ 是 $n \times n$ Hermite 矩阵,令

$$R(\boldsymbol{x}) = \frac{(\boldsymbol{Hx}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}, \quad \boldsymbol{x} \in \mathbf{C}^n, \ \boldsymbol{x} \neq \mathbf{0} \tag{7.3.1}$$

则 $R(\boldsymbol{x})$ 是实值函数,称 $R(\boldsymbol{x})$ 为 $\boldsymbol{H}$ 的 Rayleigh(瑞利)商.

关于 Rayleigh 商,我们有如下性质.

**性质 7.3.1** $\underset{\boldsymbol{x}\neq\mathbf{0}}{R(\boldsymbol{x})} = \underset{\|\boldsymbol{x}\|_2=1}{R}(\boldsymbol{x})$.

**性质 7.3.2** $R(\boldsymbol{x})$ 是连续的实值函数.

**证**　任取 $\boldsymbol{x}=(x_1, x_2, \cdots, x_n)^{\mathrm{T}}\in\mathbf{C}^n$，$\boldsymbol{y}=(y_1, y_2, \cdots, y_n)^{\mathrm{T}}\in\mathbf{C}^n$，$\|\boldsymbol{x}\|_2=1$，$\|\boldsymbol{y}\|_2=1$，$\boldsymbol{H}=(a_{ij})\in\mathbf{C}^{n\times n}$，则

$$
\begin{aligned}
|R(\boldsymbol{x})-R(\boldsymbol{y})| &= \left|\sum_{i,j=1}^{n}a_{ij}(\bar{x}_i x_j-\bar{y}_i y_j)\right| \leqslant \max_{i,j}|a_{ij}|\sum_{i,j=1}^{n}|\bar{x}_i x_j-\bar{y}_i y_j| \\
&= \frac{1}{n}\|\boldsymbol{H}\|_{m_\infty}\sum_{i,j=1}^{n}|\bar{x}_i x_j-\bar{y}_i y_j+\bar{x}_i y_j-\bar{x}_i y_j| \\
&\leqslant \frac{1}{n}\|\boldsymbol{H}\|_{m_\infty}\left(\sum_{i,j=1}^{n}|x_i||x_j-y_j|+\sum_{i,j=1}^{n}|y_j||x_i-y_i|\right) \\
&= \frac{1}{n}\|\boldsymbol{H}\|_{m_\infty}\left(2n\sum_{j=1}^{n}|x_j-y_j|^2\right)^{\frac{1}{2}} \\
&= 2\|\boldsymbol{H}\|_{m_\infty}\left(\sum_{j=1}^{n}|x_j-y_j|^2\right)^{\frac{1}{2}}
\end{aligned}
$$

于是由上式知，当 $\boldsymbol{y}\to\boldsymbol{x}$ 时

$$|R(\boldsymbol{x})-R(\boldsymbol{y})|\to 0$$

□

根据性质 7.3.2，因为 $R(\boldsymbol{x})$是复空间 $\mathbf{C}^n$中的单位闭球面 $S=\{z\in\mathbf{C}^n \mid \|z\|_2=1\}$ 上的连续实值函数，所以，$R(\boldsymbol{x})$必有最大值和最小值.

以下设 $\lambda_1\leqslant\lambda_2\leqslant\cdots\leqslant\lambda_n$ 是 $n$ 阶 Hermite 矩阵的特征值，$\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$是分别属于 $\lambda_1, \lambda_2, \cdots, \lambda_n$的单位特征向量. 令 $V_p=\mathrm{span}\{\boldsymbol{x}_p, \boldsymbol{x}_{p+1}, \cdots, \boldsymbol{x}_n\}$，$p=1, 2, \cdots, n$，则有向量子空间链

$$V_n\subset V_{n-1}\subset\cdots\subset V_2\subset V_1=\mathbf{C}^n$$

**定理 7.3.1**　$\lambda_1=\min\limits_{\|x\|_2=1}R(\boldsymbol{x})$，$\lambda_n=\max\limits_{\|x\|_2=1}R(\boldsymbol{x})$.

**证**　任取 $\boldsymbol{x}\in\mathbf{C}^n$，$\|\boldsymbol{x}\|_2=1$，则因 $\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n$是 $\mathbf{C}^n$的基，从而

$$\boldsymbol{x}=k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_n\boldsymbol{x}_n$$

其中
$$|k_1|^2+|k_2|^2+\cdots+|k_n|^2=1$$

令 $\boldsymbol{U}=(\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$，则

$$\boldsymbol{U}^{\mathrm{H}}\boldsymbol{H}\boldsymbol{U}=\begin{pmatrix}\lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n\end{pmatrix}=\boldsymbol{\Lambda} \tag{7.3.2}$$

令 $\boldsymbol{y}=\boldsymbol{U}^{\mathrm{H}}\boldsymbol{x}$,那么 $\|\boldsymbol{y}\|_2=1$,且

$$
\begin{aligned}
R(\boldsymbol{x}) &= (\boldsymbol{Hx},\boldsymbol{x}) = (\boldsymbol{HUy},\boldsymbol{Uy}) = \boldsymbol{y}^{\mathrm{H}}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{HU})\boldsymbol{y} = \boldsymbol{y}^{\mathrm{H}}\boldsymbol{\Lambda y} \\
&= \lambda_1 |y_1|^2 + \lambda_2 |y_2|^2 + \cdots + \lambda_n |y_n|^2
\end{aligned}
$$

其中 $\boldsymbol{y}=(y_1, y_2, \cdots, y_n)^{\mathrm{T}}$. 由上式即知

$$\lambda_1 \leqslant R(\boldsymbol{x}) \leqslant \lambda_n$$

若取 $\boldsymbol{x}=\boldsymbol{x}_1$,则

$$R(\boldsymbol{x}_1) = (\boldsymbol{Hx}_1, \boldsymbol{x}_1) = (\lambda_1\boldsymbol{x}_1, \boldsymbol{x}_1) = \lambda_1(\boldsymbol{x}_1, \boldsymbol{x}_1) = \lambda_1$$

若取 $\boldsymbol{x}=\boldsymbol{x}_n$,就有

$$R(\boldsymbol{x}_n) = \boldsymbol{\lambda}_n$$

所以

$$\lambda_1 = \min_{\|\boldsymbol{x}\|_2=1} R(\boldsymbol{x}) = R(\boldsymbol{x}_1), \qquad \lambda_n = \max_{\|\boldsymbol{x}\|_2=1} R(\boldsymbol{x}) = R(\boldsymbol{x}_n)$$

结合性质 7.3.1 与性质 7.3.2,上式可以等价地写成

$$\lambda_1 = \min_{0\neq\boldsymbol{x}\in V_1} R(\boldsymbol{x}), \qquad \lambda_n = \max_{0\neq\boldsymbol{x}\in V_1} R(\boldsymbol{x})$$

将定理 7.3.1 的结论推广,我们还可以得到更一般的结果.

**定理 7.3.2**

$$\lambda_i = \min_{0\neq\boldsymbol{x}\in V_i} R(\boldsymbol{x}) \tag{7.3.3}$$

$$\lambda_i = \max_{0\neq\boldsymbol{x}\in V_{i+1}^{\perp}} R(\boldsymbol{x}), \quad i=2, \cdots, n-1 \tag{7.3.4}$$

**证** 我们只证式(7.3.3),式(7.3.4)的证明与此类似.

任取 $\boldsymbol{x}\in V_i$, $\boldsymbol{x}\neq\boldsymbol{0}$,则由 $V_i$ 的定义可知,存在 $k_i, k_{i+1}, \cdots, k_n \in \mathbf{C}$ 使得 $\boldsymbol{x}=\sum_{j=i}^{n} k_j\boldsymbol{x}_j$,其中 $k_i, k_{i+1}, \cdots, k_n$不全为 0. 令

$$\boldsymbol{y} = (0, \cdots, 0, k_i, k_{i+1}, \cdots, k_n)^{\mathrm{T}}$$

显然有 $\boldsymbol{x}=\boldsymbol{Uy}$. 故

$$(\boldsymbol{Hx},\boldsymbol{x}) = (\boldsymbol{HUy},\boldsymbol{Uy}) = \boldsymbol{y}^{\mathrm{H}}(\boldsymbol{U}^{\mathrm{H}}\boldsymbol{HU})\boldsymbol{y} = \boldsymbol{y}^{\mathrm{H}}\boldsymbol{\Lambda y} = \sum_{j=i}^{n}\lambda_j |k_j|^2$$

$$(\boldsymbol{x},\boldsymbol{x}) = \sum_{j=i}^{n} |k_j|^2$$

所以

$$R(\boldsymbol{x})=\frac{(\boldsymbol{Hx},\boldsymbol{x})}{(\boldsymbol{x},\boldsymbol{x})}=\frac{\lambda_i|k_i|^2+\lambda_{i+1}|k_{i+1}|^2+\cdots+\lambda_n|k_n|^2}{|k_i|^2+|k_{i+1}|^2+\cdots+|k_n|^2}\geqslant\lambda_i$$

即 $R(\boldsymbol{x})$在 $V_i$ 上是有下界的. 但 $\boldsymbol{x}=\boldsymbol{x}_i$ 时，$R(\boldsymbol{x}_i)=\lambda_i$，$\boldsymbol{x}_i\in V_i$，所以 $R(\boldsymbol{x})$在 $V_i$ 上有最小值 $\lambda_i$. □

式(7.3.3)、式(7.3.4)表明，$\lambda_i$ 可以由 Rayleigh 商 $R(\boldsymbol{x})$在 $V_i$ 上的局部极值求出. 然而定理 7.3.2 在应用上是困难的，这是因为我们必须先求出特征向量 $\boldsymbol{x}_i$，$\boldsymbol{x}_{i+1}$，…，$\boldsymbol{x}_n$. 为此，我们进行必要的改进.

**引理 7.3.1**　设 $W_j(1\leqslant j\leqslant n)$ 是 $\mathbf{C}^n$的任意一个 $(n-j+1)$ 维子空间，则

$$\min_{0\neq x\in W_j}R(\boldsymbol{x})\leqslant\lambda_j \tag{7.3.5}$$

$$\max_{0\neq x\in W_j}R(\boldsymbol{x})\leqslant\lambda_{n-j+1} \tag{7.3.6}$$

**证**　令 $\hat{W}_j=\mathrm{Span}\{\boldsymbol{x}_1,\boldsymbol{x}_2,\cdots,\boldsymbol{x}_j\}$，$j=1,2,\cdots,n$. 对取定的 $j$，由于

$$\dim W_j+\dim\hat{W}_j=(n-j+1)+j=n+1>n$$

则必有非零向量 $\boldsymbol{x}_0\in W_j\cap\hat{W}_j$. 由 $\boldsymbol{x}_0\in\hat{W}_j$ 知 $\boldsymbol{x}_0=\sum_{i=1}^{j}k_i\boldsymbol{x}_i$，于是有

$$(\boldsymbol{Hx}_0,\boldsymbol{x}_0)=\left(\sum_{i=1}^{j}k_i\lambda_i\boldsymbol{x}_i,\sum_{i=1}^{j}k_i\boldsymbol{x}_j\right)=\sum_{i=1}^{j}\lambda_i|k_i|^2$$

显然有

$$\lambda_1=\sum_{i=1}^{j}|k_i|^2\leqslant(\boldsymbol{Hx}_0,\boldsymbol{x}_0)\leqslant\lambda_j\sum_{i=1}^{j}|k_i|^2$$

用 $\sum_{i=1}^{j}|k_i|^2=(\boldsymbol{x}_0,\boldsymbol{x}_0)$ 除上式得

$$\lambda_1\leqslant R(\boldsymbol{x}_0)\leqslant\lambda_j$$

上式表明

$$\min_{0\neq x\in W_j}R(\boldsymbol{x})\leqslant R(\boldsymbol{x}_0)\leqslant\lambda_j$$ □

类似地可以证式(7.3.6).

**定理 7.3.3**　(Courant-Fischer 定理)

$$\lambda_j=\max_{W_j}\min_{0\neq x\in W_j}R(\boldsymbol{x}) \tag{7.3.7}$$

$$\lambda_{n-j+1}=\min_{W_j}\max_{0\neq x\in W_j}R(\boldsymbol{x}) \tag{7.3.8}$$

定理 7.3.3 的意思是指，若对每个 $(n-j+1)$ 维子空间 $W_j$，在 $W_j-\{\mathbf{0}\}$ 中取极小值，然后对由所有的 $n-j+1$ 维子空间得到的极小值求出极大值，则该极大值就是 $\lambda_j$. 式(7.3.7)和式(7.3.8)也称为特征值的极大-极小原理和极小-极大原理.

**证** 根据式(7.3.5)、式(7.3.6)，为证式(7.3.7)，只需证明存在一个 $n-j+1$ 维子空间 $W_j$，使得 $\min\limits_{0\neq x\in W_j} R(\boldsymbol{x})=\lambda_j$. 令 $W_j=\mathrm{Span}\{\boldsymbol{x}_j,\ \boldsymbol{x}_{j+1},\ \cdots,\ \boldsymbol{x}_n\}$，则由式(7.3.5)知

$$\min_{0\neq x\in W_j} R(\boldsymbol{x})\leqslant\lambda_j$$

然而由 $W_j$ 的定义立即有 $W_j=V_j$. 于是由定理 7.3.2 得

$$\min_{0\neq x\in W_j} R(\boldsymbol{x})=\min_{0\neq x\in V_j} R(\boldsymbol{x})=\lambda_j$$

这就证明了式(7.3.7)，式(7.3.8)的证明与此类似. □

**例 7.3.1** 设 $n$ 阶正规矩阵 $\boldsymbol{A}$ 的特征值是 $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$，且 $|\lambda_1|\leqslant|\lambda_2|\leqslant\cdots\leqslant|\lambda_n|$. 证明：

$$|\lambda_n|=\|\boldsymbol{A}\|_2$$

**证** 因 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}$ 是 Hermite 矩阵，$\|\boldsymbol{A}\|_2=\sqrt{\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\text{ 的最大特征值}}$. 根据定理 7.3.1 得

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\text{ 的最大特征值}=\max_{\|x\|_2=1}(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x},\ \boldsymbol{x})$$

设 $\boldsymbol{x}_1,\ \boldsymbol{x}_2,\ \cdots,\ \boldsymbol{x}_n$ 是矩阵 $\boldsymbol{A}$ 的分别属于特征值 $\lambda_1,\ \lambda_2,\ \cdots,\ \lambda_n$ 的单位正交特征向量，则对任意的 $\boldsymbol{x}\in\mathbf{C}^n$，$\|\boldsymbol{x}\|_2=1$，存在常数 $k_1,\ k_2,\ \cdots,\ k_n$ 使得

$$\boldsymbol{x}=k_1\boldsymbol{x}_1+k_2\boldsymbol{x}_2+\cdots+k_n\boldsymbol{x}_n$$

此时

$$(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x},\ \boldsymbol{x})=(\boldsymbol{A}\boldsymbol{x},\ \boldsymbol{A}\boldsymbol{x})=\left(\sum_{i=1}^{n}k_i\lambda_i\boldsymbol{x}_i,\ \sum_{j=1}^{n}k_j\lambda_j\boldsymbol{x}_j\right)$$

$$=\sum_{i=1}^{n}|\lambda_i|^2|k_i|^2\leqslant|\lambda_n|^2\sum_{i=1}^{n}|k_i|^2=|\lambda_n|^2$$

又

$$(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}_n,\ \boldsymbol{x}_n)=(\boldsymbol{A}\boldsymbol{x}_n,\ \boldsymbol{A}\boldsymbol{x}_n)=(\lambda_n\boldsymbol{x}_n,\ \lambda_n\boldsymbol{x}_n)=|\lambda_n|^2$$

故

$$\max_{\|x\|_2=1} R(x) = \max_{\|x\|_2=1} (A^H A x, x) = |\lambda_n|^2$$

所以

$$|\lambda_n| = \|A\|_2$$

## 7.4 正矩阵

元素都是非负实数的矩阵称为非负矩阵，这类矩阵在数理经济学、概率论、弹性系统微振动理论等许多领域都有重要的作用，其基本特征已被认为是矩阵论的经典内容之一. 本节将介绍非负矩阵的一类重要的特例——正矩阵的基本性质.

**定义 7.4.1** 设矩阵 $A = (a_{ij}) \in \mathbf{R}^{m\times n}$，若 $A$ 中所有元素都是非负实数，则称 $A$ 为非负矩阵，记为 $A \geqslant 0$；若 $A$ 中所有元素都是正数，则称 $A$ 为正矩阵，记为 $A > 0$.

设 $A, B \in \mathbf{R}^{m\times n}$，若 $A - B \geqslant 0$，则记为 $A \geqslant B$；若 $A - B > 0$，则记为 $A > B$.

Perron(佩龙)在 1907 年建立了正矩阵的特征值与特征向量的重要性质.

**定理 7.4.1** 设 $A \in \mathbf{R}^{n\times n}$ 为正矩阵，且 $\rho(A)$ 为其谱半径，则

(i) $\rho(A)$ 为 $A$ 的正特征值，它对应着一个正的特征向量；

(ii) $A$ 的任何其他特征值 $\lambda$，都有 $|\lambda| \leqslant \rho(A)$；

(iii) $\rho(A)$ 为 $A$ 的单特征值.

**证** (i) 设 $\mu$ 是 $A$ 的特征值，满足 $|\mu| = \rho(A)$，$x = (x_1, x_2, \cdots, x_n)^T$ 为属于 $\mu$ 的 $A$ 的特征向量，则

$$Ax = \mu x$$

即
$$\mu x_i = \sum_{j=1}^{n} a_{ij} x_j, \quad i = 1, 2, \cdots, n$$

由此，有

$$\rho(A)|x_i| = |\mu x_i| \leqslant \sum_{j=1}^{n} a_{ij} |x_j|, \quad i = 1, 2, \cdots, n$$

令 $y = (|x_1|, \cdots, |x_n|)^T$，则有

$$\rho(A) y \leqslant A y$$

亦即

$$(A - \rho(A)E) y \geqslant 0$$

下面证明等号成立. 用反证法.

设 $(\boldsymbol{A}-\rho(\boldsymbol{A})\boldsymbol{E})\boldsymbol{y}=\boldsymbol{z}\neq 0$. 因为 $\boldsymbol{A}>0$，所以

$$\boldsymbol{A}\boldsymbol{z}>0,\qquad \boldsymbol{A}\boldsymbol{y}>0$$

因此存在 $\varepsilon>0$，使得

$$\boldsymbol{A}\boldsymbol{z}>\varepsilon\boldsymbol{A}\boldsymbol{y}$$

又因

$$\boldsymbol{A}\boldsymbol{z}=\boldsymbol{A}(\boldsymbol{A}-\rho(\boldsymbol{A})\boldsymbol{E})\boldsymbol{y}=\boldsymbol{A}^2\boldsymbol{y}-\rho(\boldsymbol{A})\boldsymbol{A}\boldsymbol{y}$$

所以

$$\boldsymbol{A}^2\boldsymbol{y}\geqslant(\varepsilon+\rho(\boldsymbol{A}))\boldsymbol{A}\boldsymbol{y}$$

令 $\boldsymbol{B}=[\varepsilon+\rho(\boldsymbol{A})]^{-1}\boldsymbol{A}$，则有

$$\boldsymbol{B}\boldsymbol{A}\boldsymbol{y}\geqslant\boldsymbol{A}\boldsymbol{y}$$

由于 $\boldsymbol{B}>0$，故由上式逐步可得

$$\boldsymbol{B}^n\boldsymbol{A}\boldsymbol{y}\geqslant\boldsymbol{A}\boldsymbol{y},\quad n=1,2,\cdots \tag{7.4.1}$$

因为

$$\rho(\boldsymbol{B})=[\varepsilon+\rho(\boldsymbol{A})]^{-1}\rho(\boldsymbol{A})<1$$

所以

$$\lim_{n\to\infty}\boldsymbol{B}^n=0$$

在式(7.4.1)取极限，得到 $\boldsymbol{A}\boldsymbol{y}\leqslant 0$，这与 $\boldsymbol{A}\boldsymbol{y}>0$ 矛盾. 从而

$$\boldsymbol{A}\boldsymbol{y}=\rho(\boldsymbol{A})\boldsymbol{y}$$

亦即 $\rho(\boldsymbol{A})$ 为 $\boldsymbol{A}$ 的正特征值，且 $\boldsymbol{y}$ 为属于 $\rho(\boldsymbol{A})$ 的一个正的特征向量.

(ii) 设 $\lambda$ 是 $\boldsymbol{A}$ 的特征值，满足 $|\lambda|=\rho(\boldsymbol{A})$，下面证明 $\lambda=\rho(\boldsymbol{A})$. 设 $\boldsymbol{u}=(u_1,\cdots,u_n)^{\mathrm{T}}$ 为属于 $\lambda$ 的特征向量，则

$$\boldsymbol{A}\boldsymbol{u}=\lambda\boldsymbol{u}$$

即

$$\lambda u_i=\sum_{j=1}^n a_{ij}u_j,\quad i=1,\cdots,n \tag{7.4.2}$$

令 $\boldsymbol{v}=(|u_1|,\cdots,|u_n|)^{\mathrm{T}}$，类似(1)的方法可以得到 $\boldsymbol{A}\boldsymbol{v}=\rho(\boldsymbol{A})\boldsymbol{v}$，即

$$\rho(\boldsymbol{A}) \mid u_i \mid = \sum_{j=1}^{n} a_{ij} \mid u_j \mid, \quad i = 1, \cdots, n \tag{7.4.3}$$

在式(7.4.2)两边取绝对值,根据(7.4.3)得

$$\left| \sum_{j=1}^{n} a_{ij} u_j \right| = \sum_{j=1}^{n} a_{ij} \mid u_j \mid, \quad i = 1, \cdots, n$$

上式表明,$u_j (j = 1, \cdots, n)$ 有相同的幅角 $\theta$,即

$$u_j = \mid u_j \mid \mathrm{e}^{i\theta}, \quad j = 1, \cdots, n$$

于是

$$\boldsymbol{u} = \boldsymbol{v}\mathrm{e}^{i\theta}$$

从而 $\boldsymbol{u}$ 也是属于 $\rho(\boldsymbol{A})$ 的 $\boldsymbol{A}$ 的特征向量. 而 $u$ 是属于 $\lambda$ 的 $\boldsymbol{A}$ 的特征向量,因此

$$\lambda = \rho(\boldsymbol{A})$$

(iii) 令 $\boldsymbol{B} = \rho(\boldsymbol{A})^{-1}\boldsymbol{A} = (b_{ij})$,则 $\boldsymbol{B} > 0$,且 $\rho(\boldsymbol{B}) = 1$. 下面证明 1 是 $\boldsymbol{B}$ 的单特征值. 为此,只需证明 $\boldsymbol{B}$ 的 Jordan 标准形中对应于特征值 1 有且只有一个一阶的 Jordan 块. 由(i),有向量

$$\boldsymbol{y} = (y_1, \cdots, y_n)^{\mathrm{T}} > 0$$

使得

$$\boldsymbol{B}\boldsymbol{y} = \boldsymbol{y}$$

从而

$$\boldsymbol{B}^k \boldsymbol{y} = \boldsymbol{y}, \quad k = 1, 2, \cdots$$

令 $y_M$, $y_m$ 分别为 $\boldsymbol{y}$ 的最大和最小的分量,则有

$$y_M \geqslant y_i = \sum_{l=1}^{n} b_{il}^{(k)} y_l \geqslant b_{ij}^{(k)} y_j \geqslant y_m$$

其中 $b_{ij}^{(k)}$ 为矩阵 $\boldsymbol{B}^k$ 的 $(i, j)$ 位置上的元素. 从而

$$b_{ij}^{(k)} \leqslant \frac{y_M}{y_m}$$

这表明,对于所有的 $k$,$b_{ij}^{(k)}$ 有界. 假若 $\boldsymbol{B}$ 的 Jordan 标准形 $\boldsymbol{J}$ 中有一个对应于特征值 1 的 Jordan 块的阶数大于 1,不妨设为 2,则存在可逆阵 $\boldsymbol{P}$,使得

$$\boldsymbol{P}^{-1}\boldsymbol{B}^k\boldsymbol{P} = \boldsymbol{J}^k = \begin{bmatrix} 1 & k & & & \\ & 1 & & & \\ & & \boldsymbol{J}_1^k(\lambda_1) & & \\ & & & \ddots & \\ & & & & \boldsymbol{J}_m^k(\lambda_m) \end{bmatrix}$$

其中 $J_i(\lambda_i)$ 为 Jordan 块. 因为 $B^k$ 中元素 $b_{ij}^{(k)}$ 对所有的 $k$ 有界，所以 $J^k$ 中元素对所有的 $k$ 有界，但是上式结果与此矛盾，因此 $J$ 对应于特征值 1 的 Jordan 块只能是一阶的. 下面证明这种 Jordan 块只有一个. 设 $B$ 的 Jordan 标准形为

$$J = \begin{pmatrix} E_r & & & \\ & J_1(\lambda_1) & & \\ & & \ddots & \\ & & & J_l(\lambda_l) \end{pmatrix}$$

其中，$E_r$ 为 $r$ 阶单位矩阵；$|\lambda_i| < 1$，$i = 1, \cdots, l$. 所以 $(E-J)x = 0$ 的解空间的维数为 $r$，从而 $(E-B)x = 0$ 的解空间的维数也为 $r$. 如果 $r > 1$，则特征值 1 至少有一个和 $y$ 线性无关的特征向量

$$z = (z_1, \cdots, z_n)^{\mathrm{T}}$$

因为 $1 = \rho(B)$，所以由(ii)和(i)知 $z$ 为正向量. 令

$$\tau = \max\left\{\frac{z_i}{y_i};\ i = 1, \cdots, n\right\} = \frac{z_j}{y_j}$$

则有 $\tau y \geqslant z$，且由于 $y$ 和 $z$ 线性无关，不能取等式，由此，有

$$B(\tau y - z) > 0$$

又因为

$$By = y, \qquad Bz = z$$

所以

$$\tau y - z > 0$$

于是

$$\tau > \frac{z_j}{y_j}$$

这与 $\tau$ 的定义矛盾，故

$$r = 1$$

□

## 7.5 非负矩阵

Perron 定理显示了正矩阵具有很好的谱性质. 1912 年，Frobenius 将 Perron

定理推广到不可约的非负矩阵上.我们先介绍不可约矩阵的概念.

一个方阵 $\boldsymbol{P}$,若它的每一行和每一列都只有某个元素为1,其余的元素为0,则矩阵 $\boldsymbol{P}$ 称为一个置换矩阵.

**定义 7.5.1** 矩阵 $\boldsymbol{A}=(a_{ij})\in\mathbf{R}^{n\times n}$,如果存在 $n$ 阶置换矩阵 $\boldsymbol{P}$,使得

$$\boldsymbol{PAP}^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{A}_{11} & \boldsymbol{0}\\ \boldsymbol{A}_{21} & \boldsymbol{A}_{22}\end{bmatrix}$$

其中,$\boldsymbol{A}_{11}$ 为 $k$ 阶方阵;$\boldsymbol{A}_{22}$ 为 $n-k$ 阶方阵($1\leqslant k\leqslant n-1$),则称 $\boldsymbol{A}$ 为可约的;否则,称 $\boldsymbol{A}$ 为不可约的.

据此定义,一阶方阵和正矩阵都是不可约的.

**定理 7.5.1** 设 $\boldsymbol{A}\in\mathbf{R}^{n\times n}$ 为非负不可约矩阵,则

(i) $\rho(\boldsymbol{A})$为 $\boldsymbol{A}$ 的正特征值,其对应着一个正的特征向量;

(ii) $\boldsymbol{A}$ 的任何其他特征值 $\lambda$,都有 $|\lambda|\leqslant\rho(\boldsymbol{A})$;

(iii) $\rho(\boldsymbol{A})$为 $\boldsymbol{A}$ 的单特征值;

(iv) 当 $\boldsymbol{A}$ 的任一元素增加时,$\rho(\boldsymbol{A})$增加.

这个定理的证明非常复杂,故略去.

对于一般的非负矩阵,上面的定理虽然不成立,但是有下面的结果(证略):

**定理 7.5.2** 设 $\boldsymbol{A}\in\mathbf{R}^{n\times n}$ 为非负矩阵,则 $\boldsymbol{A}$ 必有一非负特征值 $\lambda$,而 $\boldsymbol{A}$ 的所有特征值的模都不超过 $\lambda$,且特征值 $\lambda$ 对应着一非负特征向量.

此外,非负矩阵还有一个比较重要的性质,即若 $0\leqslant\boldsymbol{B}\leqslant\boldsymbol{A}$,则 $\rho(\boldsymbol{B})\leqslant\rho(\boldsymbol{A})$.关于非负矩阵的谱半径,有如下定理.

**定理 7.5.3** 设 $\boldsymbol{A}\in\mathbf{R}^{n\times n}$ 为非负矩阵,则

(i) 若 $\boldsymbol{A}$ 的每一行元素之和为常数,则 $\rho(\boldsymbol{A})=\|\boldsymbol{A}\|_\infty$;

(ii) 若 $\boldsymbol{A}$ 的每一列元素之和为常数,则 $\rho(\boldsymbol{A})=\|\boldsymbol{A}\|_1$;

(iii) $\min\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^{n}a_{ij}\leqslant\rho(\boldsymbol{A})\leqslant\max\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^{n}a_{ij}$;

(iv) $\min\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^{n}a_{ij}\leqslant\rho(\boldsymbol{A})\leqslant\max\limits_{1\leqslant j\leqslant n}\sum\limits_{i=1}^{n}a_{ij}$.

**证** 首先,因对于 $\boldsymbol{A}$ 的任何特征值 $\lambda$ 和任何相容的矩阵范数 $\|\cdot\|$,有 $\lambda\leqslant\|\boldsymbol{A}\|$,所以 $\rho(\boldsymbol{A})\leqslant\|\boldsymbol{A}\|_\infty$;其次,由 $\boldsymbol{Au}=\|\boldsymbol{A}\|_\infty\boldsymbol{u}$, $\boldsymbol{u}=(1,\cdots,1)^{\mathrm{T}}$ 知 $\|\boldsymbol{A}\|_\infty$ 为 $\boldsymbol{A}$ 的特征值,所以有 $\|\boldsymbol{A}\|_\infty\leqslant\rho(\boldsymbol{A})$. 这就证明了(i).

同样的讨论应用于 $\boldsymbol{A}^{\mathrm{T}}$ 可以得到(ii).

设 $m=\min\limits_{1\leqslant i\leqslant n}\sum\limits_{j=1}^{n}a_{ij}$.构造矩阵 $\boldsymbol{B}=(b_{ij})$:若 $m=0$,令 $\boldsymbol{B}=\boldsymbol{0}$;若 $m>0$,令 $b_{ij}=ma_{ij}\left(\sum\limits_{j=1}^{n}a_{ij}\right)^{-1}$,则有 $0\leqslant\boldsymbol{B}\leqslant\boldsymbol{A}$,且 $\sum\limits_{j=1}^{n}b_{ij}=m$, $i=1,\cdots,n$. 由(i)知 $\rho(\boldsymbol{B})=m$.

另一方面，由于$\rho(\boldsymbol{B})\leqslant\rho(\boldsymbol{A})$，所以$m\leqslant\rho(\boldsymbol{A})\leqslant\|\boldsymbol{A}\|_\infty$，亦即(iii)成立. 对矩阵$\boldsymbol{A}^{\mathrm{T}}$应用(iii)可得到(iv). □

最后，注意到对于一个给定的矩阵，直接根据定义 7.5.1 判断其是否可约是比较困难的，特别是矩阵的阶数比较大的时候. 下面我们给出非负矩阵是否可约的一个判别条件.

**定理 7.5.4** $n$ 阶非负矩阵$\boldsymbol{A}$ 为不可约的充要条件是存在正整数 $s\leqslant n-1$，使得

$$(\boldsymbol{E}+\boldsymbol{A})^s>0$$

**证** 必要性. 设 $A$ 为 $n$ 阶不可约非负矩阵. 只需证明对任意向量 $\boldsymbol{y}\geqslant 0$($\boldsymbol{y}\neq 0$) 都有

$$(\boldsymbol{E}+\boldsymbol{A})^{n-1}\boldsymbol{y}>0$$

成立即可. 对任意的 $\boldsymbol{x}>0$，$\boldsymbol{x}\neq 0$，因为 $\boldsymbol{E}+\boldsymbol{A}$ 的对角线元素非零，所以 $\boldsymbol{y}=(\boldsymbol{E}+\boldsymbol{A})\boldsymbol{x}$ 中零坐标的个数小于向量 $\boldsymbol{x}$ 中零坐标的个数. 假设相反，那么 $\boldsymbol{x}$ 与 $\boldsymbol{y}$ 有相同的零坐标个数. 不失一般，可设

$$\boldsymbol{x}=\begin{pmatrix}\boldsymbol{x}_1\\ \boldsymbol{0}\end{pmatrix},\qquad \boldsymbol{y}=\begin{pmatrix}\boldsymbol{y}_1\\ \boldsymbol{0}\end{pmatrix}$$

其中 $\boldsymbol{x}_1$，$\boldsymbol{y}_1$为 $r$ 维正向量. 记 $\boldsymbol{A}$ 为

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{A}_1 & \boldsymbol{A}_2\\ \boldsymbol{A}_3 & \boldsymbol{A}_4\end{pmatrix}$$

其中 $\boldsymbol{A}_1$为 $r$ 阶方阵，则

$$\boldsymbol{y}=\begin{pmatrix}\boldsymbol{y}_1\\ \boldsymbol{0}\end{pmatrix}=(\boldsymbol{E}+\boldsymbol{A})\boldsymbol{x}=\begin{pmatrix}(\boldsymbol{E}_r+\boldsymbol{A}_1)\boldsymbol{x}_1\\ \boldsymbol{A}_3\boldsymbol{x}_1\end{pmatrix}$$

由此 $\boldsymbol{A}_3\boldsymbol{x}_1=\boldsymbol{0}$. 因为 $\boldsymbol{x}_1$ 为正向量，所以 $\boldsymbol{A}_3=\boldsymbol{0}$. 这与 $\boldsymbol{A}$ 为不可约矩阵矛盾，所以 $\boldsymbol{y}$ 中零坐标的个数小于向量 $\boldsymbol{x}$ 中零坐标的个数. 上述结果表明，每用 $\boldsymbol{E}+\boldsymbol{A}$ 对向量 $\boldsymbol{y}$ 左乘一次，其零坐标个数至少减小一个，从而得到 $(\boldsymbol{E}+\boldsymbol{A})^{n-1}\boldsymbol{y}>0$.

充分性. 设有 $(\boldsymbol{E}+\boldsymbol{A})^s>0$. 若 $\boldsymbol{A}$ 为可约的，则存在置换矩阵 $\boldsymbol{P}$，使得

$$\boldsymbol{PAP}^{\mathrm{T}}=\begin{pmatrix}\boldsymbol{A}_1 & \boldsymbol{0}\\ \boldsymbol{A}_2 & \boldsymbol{A}_3\end{pmatrix}$$

其中 $\boldsymbol{A}_1$，$\boldsymbol{A}_3$为方阵(设阶数为 $r$ 和 $t$). 于是

$$P(E+A)^s P^{\mathrm{T}} = \begin{pmatrix} (E_r+A_1)^s & 0 \\ B_s & (E_t+A_3)^s \end{pmatrix}$$

因为 $P$ 为置换矩阵，上式表明 $(E+A)^s$ 中有零元素，这与 $(E+A)^s>0$ 矛盾，所以 $A$ 不可约. □

**例 7.5.1**　非负矩阵 $A=\begin{pmatrix}1&1&0\\1&1&1\\0&1&1\end{pmatrix}$ 是不可约的. 事实上，当 $s=3-1=2$ 时，有

$$(E+A)^2=\begin{pmatrix}5&4&1\\4&5&4\\1&4&5\end{pmatrix}>0$$

## 7.6　随机矩阵

**定义 7.6.1**　非负矩阵 $A=(a_{ij})\in \mathbf{R}^{n\times n}$ 称为一个随机矩阵，且 $A$ 的每一行上的元素之和等于 1.

随机矩阵有重要的应用价值，它在 Markov(马尔可夫)过程理论中起着重要的作用，也常常出现在经济学和运筹学的各种问题中. 随机矩阵 $A$ 的每一行可以看成有 $n$ 个点的样本空间上的离散概率分布. 据定义知，随机矩阵 $A$ 有特征值 1，且对应于 1 的正特征向量为 $u=(1,\cdots,1)^{\mathrm{T}}$；反之，如果 $n$ 阶非负矩阵 $A$ 有特征值 1，且对应于 1 的特征向量为 $u=(1,\cdots,1)^{\mathrm{T}}$，则 $A$ 为随机矩阵. 于是有如下定理.

**定理 7.6.1**　$n$ 阶非负矩阵 $A$ 是随机矩阵的充要条件是 $u=(1,\cdots,1)^{\mathrm{T}}$ 为 $A$ 对应于特征值 1 的特征向量.

由定理 7.5.3 知，随机矩阵 $A$ 的谱半径 $\rho(A)=1$.

下面定理揭示了具有正谱半径与对应的正特征向量的非负矩阵与随机矩阵之间存在密切的关系.

**定理 7.6.2**　设非负矩阵 $A$ 的谱半径 $\rho(A)>0$，且有 $x=(x_1,\cdots,x_n)^{\mathrm{T}}>0$，使得 $Ax=\rho(A)x$，则存在对角元为正数的对角矩阵 $D$ 使得 $\rho(A)^{-1}D^{-1}AD$ 为随机矩阵.

**证**　由条件，得

$$Ax=\rho(A)x,\qquad x=(x_1,\cdots,x_n)^{\mathrm{T}}>0$$

令

$$D = \text{diag}(x_1, \cdots, x_n), \qquad P = \rho(A)^{-1} D^{-1} AD$$

则 $P$ 为非负矩阵，且满足

$$Pu = PD^{-1}x = \rho(A)^{-1} D^{-1} Ax = D^{-1}x = u$$

其中 $u = (1, \cdots, 1)^T$. 从而，由定理 7.5.4 知 $P$ 为随机矩阵. □

实际应用中常常要考虑随机矩阵 $A$ 的幂序列$\{A^n\}$的收敛性.

**定理 7.6.3** 设 $A$ 为 $n$ 阶随机矩阵，则 $\lim\limits_{m\to\infty} A^m$ 存在的充分必要条件是 $A$ 的不等于 1 的特征值的模均小于 1.

**证** 可以证明：如果 $A$ 为 $n$ 阶随机矩阵，则 $A$ 对应于特征值 1 的 Jordan 块均是一阶的，于是存在可逆阵 $P$ 使得

$$A = P\begin{pmatrix} E_r & & & \\ & J_1(\lambda_1) & & \\ & & \ddots & \\ & & & J_l(\lambda_l) \end{pmatrix} P^{-1}$$

其中，$J_i(\lambda_i)$ 是 Jordan 块；$\lambda_i \neq 1$，$|\lambda_i| < 1$，$i = 1, \cdots, l$. 则对于任意正整数 $m$，有

$$A^m = P\begin{pmatrix} E_r & & & \\ & J_1^m(\lambda_1) & & \\ & & \ddots & \\ & & & J_l^m(\lambda_l) \end{pmatrix} P^{-1}$$

由此式可知，$\lim\limits_{m\to\infty} A^m$ 存在的充要条件是

$$|\lambda_i| < 1, \quad i = 1, \cdots, l$$ □

**例 7.6.1** 考虑一地区人口流动问题. 设该地区有专门人才 1 800 人，分布在 $A$，$B$，$C$ 三个单位. 每年每个单位把所有人才各分二分之一与其他两个单位交流. 今年 $A$，$B$，$C$ 三个单位各有人才分别为 200 人，600 人，1 000 人，问明年、后年的分布情况如何？很多年后各单位的人才期望值是多少？

**解** 用 $\pi_A^k$，$\pi_B^k$，$\pi_C^k$分别表示第 $k$ 年 $A$，$B$，$C$ 三个单位的人才数，则 $\boldsymbol{\pi}^k = (\pi_A^k, \pi_B^k, \pi_C^k)$ 表示一个具有三个状态的齐次 Markov 链，它的转移矩阵和初始分布向量分别为

$$\boldsymbol{P}=\begin{pmatrix}0 & \frac{1}{2} & \frac{1}{2}\\ \frac{1}{2} & 0 & \frac{1}{2}\\ \frac{1}{2} & \frac{1}{2} & 0\end{pmatrix},\qquad \boldsymbol{\pi}^0=(200,\ 600,\ 1\ 000)$$

且 $\boldsymbol{\pi}^k$ 满足

$$\boldsymbol{\pi}^k=\boldsymbol{\pi}^{k-1}\boldsymbol{P}=\boldsymbol{\pi}^0\boldsymbol{P}^k$$

从而

$$\boldsymbol{\pi}^1=(800,\ 600,\ 400)$$

$$\boldsymbol{\pi}^2=(500,\ 600,\ 700)$$

由于 $\boldsymbol{P}$ 的特征值为 $1,\ -\frac{1}{2},\ -\frac{1}{2}$，从而由定理 7.6.3 知 $\lim\limits_{m\to\infty}\boldsymbol{P}^m$ 存在. 事实上

$$\boldsymbol{P}^m=\begin{pmatrix}p_m & p_{m+1} & p_{m+1}\\ p_{m+1} & p_m & p_{m+1}\\ p_{m+1} & p_{m+1} & p_m\end{pmatrix},\qquad p_m=\frac{1}{3}\left[1+\frac{(-1)^m}{2^{m-1}}\right]$$

从而

$$\lim_{m\to\infty}\boldsymbol{P}_m=\frac{1}{3}\begin{pmatrix}1 & 1 & 1\\ 1 & 1 & 1\\ 1 & 1 & 1\end{pmatrix}$$

这表明很多年后 $A$，$B$，$C$ 三个单位的人才数为一稳态的分布向量

$$\boldsymbol{\pi}=\lim_{m\to\infty}\boldsymbol{\pi}^m=\boldsymbol{\pi}^0\lim_{m\to\infty}\boldsymbol{P}^m=(600,\ 600,\ 600)$$

此外，注意到由于 $\boldsymbol{E}+\boldsymbol{P}>0$，故 $\boldsymbol{P}$ 为不可约的非负矩阵. □

## 7.7　M 矩阵

1937 年，Ostrowski（奥斯特洛夫斯基）发现一类具有特殊构造的矩阵，其非对角元素都是非正的，即这种矩阵 $\boldsymbol{A}$ 可以表示成

$$\boldsymbol{A}=s\boldsymbol{E}-\boldsymbol{B}$$

其中 $s>0$，$\boldsymbol{B}\geqslant 0$，故这种矩阵和非负矩阵有一定的联系. 现在这种矩阵在偏微分

方程的有限差分、经济系统的投入产出分析、运筹学的线性互补问题等很多领域有重要的应用. 这里讨论重要的一种,称为 Minkowski(闵可夫斯基)矩阵,简称 M 矩阵.

**定义 7.7.1** 设

$$\boldsymbol{A}=s\boldsymbol{E}-\boldsymbol{B}$$

为 $n$ 阶实矩阵,其中 $s>0$, $\boldsymbol{B}\geqslant 0$. 若 $s\geqslant\rho(\boldsymbol{B})$,则称 $\boldsymbol{A}$ 为 M 矩阵;若 $s>\rho(\boldsymbol{B})$,则称 $\boldsymbol{A}$ 为非奇异 M 矩阵.

记 $\mathbf{Z}^{n\times n}=\{\boldsymbol{A}=(a_{ij})\in\mathbf{R}^{n\times n}\mid a_{ij}\leqslant 0,\ i\neq j,\ i,\ j=1,\ \cdots,\ n\}$. 非奇异 M 矩阵有如下特性.

**定理 7.7.1** 设 $\boldsymbol{A}\in\mathbf{Z}^{n\times n}$ 为非奇异 M 矩阵,且 $\boldsymbol{D}\in\mathbf{Z}^{n\times n}$ 满足 $\boldsymbol{A}\leqslant\boldsymbol{D}$, 则

(i) $\boldsymbol{A}^{-1}$ 与 $\boldsymbol{D}^{-1}$ 存在,且 $\boldsymbol{A}^{-1}\geqslant\boldsymbol{D}^{-1}\geqslant 0$;

(ii) $\boldsymbol{D}$ 的每一实特征值为正数;

(iii) $|\boldsymbol{D}|\geqslant|\boldsymbol{A}|>0$.

**证** (i) 设 $\boldsymbol{A}=s\boldsymbol{E}-\boldsymbol{B}$, 其中 $s>\rho(\boldsymbol{B})$, $\boldsymbol{B}\geqslant 0$. 对任意给定的实数 $w\leqslant 0$, 考虑矩阵

$$\boldsymbol{C}=\boldsymbol{A}-w\boldsymbol{E}=(s-w)\boldsymbol{E}-\boldsymbol{B}$$

由于 $s-w>\rho(\boldsymbol{B})$ ,所以 $\boldsymbol{C}$ 为非奇异 M 矩阵. 这表明非奇异 M 矩阵的每个实特征值必是正数. 由于 $\boldsymbol{D}\in\mathbf{Z}^{n\times n}$, 故存在足够小的正数 $\varepsilon$,使

$$\boldsymbol{P}=\boldsymbol{E}-\varepsilon\boldsymbol{D}\geqslant 0$$

令

$$\boldsymbol{Q}=\boldsymbol{E}-\varepsilon\boldsymbol{A}$$

由 $\boldsymbol{A}\leqslant\boldsymbol{D}$ 知

$$\boldsymbol{E}-\varepsilon\boldsymbol{A}\geqslant\boldsymbol{E}-\varepsilon\boldsymbol{D}$$

亦即

$$0\leqslant\boldsymbol{P}\leqslant\boldsymbol{Q}$$

于是,$\boldsymbol{Q}$ 为非负矩阵,其谱半径 $\rho(\boldsymbol{Q})$ 为其非负特征值,从而

$$|(1-\rho(\boldsymbol{Q}))\boldsymbol{E}-\varepsilon\boldsymbol{A}|=|\boldsymbol{Q}-\rho(\boldsymbol{Q})\boldsymbol{E}|=0$$

由此 $\frac{1}{\varepsilon}[1-\rho(\boldsymbol{Q})]$ 为 $\boldsymbol{A}$ 的实特征值. 根据上面的分析, $1-\rho(\boldsymbol{Q})>0$. 因此有

$$(\varepsilon\boldsymbol{A})^{-1}=(\boldsymbol{E}-\boldsymbol{Q})^{-1}=\boldsymbol{E}+\boldsymbol{Q}+\boldsymbol{Q}^2+\cdots$$

这说明 $\boldsymbol{A}^{-1} \geqslant 0$. 又因

$$0 \leqslant \boldsymbol{P}^k \leqslant \boldsymbol{Q}^k, \quad k = 1, 2, \cdots$$

且 $\rho(\boldsymbol{P}) \leqslant \rho(\boldsymbol{Q}) < 1$，所以

$$(\varepsilon \boldsymbol{D})^{-1} = (\boldsymbol{E} - \boldsymbol{P})^{-1} = \boldsymbol{E} + \boldsymbol{P} + \boldsymbol{P}^2 + \cdots \leqslant (\varepsilon \boldsymbol{A})^{-1}$$

从而得

$$\boldsymbol{A}^{-1} \geqslant \boldsymbol{D}^{-1} \geqslant 0$$

(ii) 取 $w \leqslant 0$，则

$$\boldsymbol{D} - w\boldsymbol{E} \geqslant \boldsymbol{A}$$

由(i)知 $\boldsymbol{D} - w\boldsymbol{E}$ 非奇异，所以 $\boldsymbol{D}$ 的所有实特征值为正数.

(iii) 由前面的证明可知，$\boldsymbol{A}$ 的所有实特征值为正数，所以只需证明：若 $\boldsymbol{A}$ 的所有实特征值为正数，且 $\boldsymbol{A} \leqslant \boldsymbol{D}$，则(iii)成立. 对矩阵的阶数 $n$ 作归纳法. 设 $\boldsymbol{A}_1$，$\boldsymbol{D}_1$ 分别是 $\boldsymbol{A}$，$\boldsymbol{D}$ 的前 $n-1$ 行、前 $n-1$ 列构成的矩阵，则 $\boldsymbol{A}_1, \boldsymbol{D}_1 \in \mathbf{Z}^{(n-1)\times(n-1)}$，且 $\boldsymbol{A}_1 \leqslant \boldsymbol{D}_1$. 设

$$\widetilde{\boldsymbol{A}} = \begin{pmatrix} \boldsymbol{A}_1 & \boldsymbol{0} \\ \boldsymbol{0} & a_{nn} \end{pmatrix}$$

则 $\boldsymbol{A} \leqslant \widetilde{\boldsymbol{A}}$，所以由(ii)得，$\widetilde{\boldsymbol{A}}$ 的所有实特征值均为正数，从而 $\boldsymbol{A}_1$ 的实特征值为正数. 由归纳假设，$|\boldsymbol{D}_1| \geqslant |\boldsymbol{A}_1| > 0$. 又由(i)，$\boldsymbol{A}^{-1} \geqslant \boldsymbol{D}^{-1} \geqslant 0$ 知 $\boldsymbol{A}^{-1}$，$\boldsymbol{D}^{-1}$ 的 $(n, n)$ 满足

$$(\boldsymbol{A}^{-1})_{nn} \geqslant (\boldsymbol{D}^{-1})_{nn} \geqslant 0$$

于是

$$\frac{|\boldsymbol{A}_1|}{|\boldsymbol{A}|} \geqslant \frac{|\boldsymbol{D}_1|}{|\boldsymbol{D}|}$$

因此有

$$|\boldsymbol{A}| > 0, \qquad |\boldsymbol{D}| > 0$$

且有

$$|\boldsymbol{D}| \geqslant |\boldsymbol{D}_1| \cdot \frac{|\boldsymbol{A}|}{|\boldsymbol{A}_1|} \geqslant |\boldsymbol{A}| > 0$$

**定义 7.7.2** 设矩阵 $\boldsymbol{A} = (a_{ij}) \in \mathbf{C}^{n\times n}$，若

$$|a_{ii}| \geqslant \sum_{j\neq i} |a_{ij}|, \quad i = 1, \cdots, n$$

则称 $\boldsymbol{A}$ 为行对角占优的；若上述不等式是严格不等的，即

$$|a_{ii}| \geqslant \sum_{j\neq i} |a_{ij}|, \quad i=1, \cdots, n$$

则称 $\boldsymbol{A}$ 为严格行对角占优的. 类似可定义列对角占优矩阵的概念.

下面的定理列举了 $\boldsymbol{Z}^{n\times n}$ 中矩阵为非奇异 M 矩阵的一些等价条件.

**定理 7.7.2** 设 $\boldsymbol{A}\in \boldsymbol{Z}^{n\times n}$，则以下各命题彼此等价：

(i) 存在 $\boldsymbol{x}\geqslant 0$，使得 $\boldsymbol{Ax}>0$；

(ii) 存在 $\boldsymbol{x}>0$，使得 $\boldsymbol{Ax}>0$；

(iii) 存在正对角矩阵 $\boldsymbol{D}$，使得 $\boldsymbol{AD}$ 为严格行对角占优矩阵，且 $\boldsymbol{AD}$ 的所有对角线元素为正数；

(iv) $\boldsymbol{A}$ 的每个实特征值为正数；

(v) $\boldsymbol{A}$ 为非奇异 M 矩阵；

(vi) 若 $\boldsymbol{B}\in \boldsymbol{Z}^{n\times n}$ 且 $\boldsymbol{B}\geqslant \boldsymbol{A}$，则 $\boldsymbol{B}$ 非奇异；

(vii) $\boldsymbol{A}$ 的任意主子阵的每个实特征值为正数；

(viii) $\boldsymbol{A}$ 的所有主子式为正数；

(ix) 对每个 $k(1\leqslant k\leqslant n)$，$\boldsymbol{A}$ 的所有 $k$ 阶主子式之和为正数；

(x) 存在 $\boldsymbol{A}$ 的一种分裂 $\boldsymbol{A}=\boldsymbol{P}-\boldsymbol{Q}$，其中 $\boldsymbol{P}^{-1}\geqslant 0$，$\boldsymbol{Q}\geqslant 0$ 且 $\rho(\boldsymbol{P}^{-1}\boldsymbol{Q})<1$；

(xi) $\boldsymbol{A}$ 非奇异且 $\boldsymbol{A}^{-1}\geqslant 0$.

**证** (i)⇒(ii) 设 $\boldsymbol{x}\geqslant 0$，满足 $\boldsymbol{Ax}>0$. 令 $\boldsymbol{x}_0=(1, \cdots, 1)^{\mathrm{T}}$，由于 $\boldsymbol{Ax}>0$，所以存在 $\varepsilon>0$ 使

$$\boldsymbol{Ax}+\varepsilon\boldsymbol{A}\boldsymbol{x}_0>0$$

从而向量 $\boldsymbol{x}+\varepsilon\boldsymbol{x}_0>0$ 满足

$$\boldsymbol{A}(\boldsymbol{x}+\varepsilon\boldsymbol{x}_0)>0$$

(ii)⇒(iii) 设 $\boldsymbol{x}=(x_1, \cdots, x_n)^{\mathrm{T}}>0$，满足 $\boldsymbol{Ax}>0$. 令 $\boldsymbol{D}=\mathrm{diag}(x_1, \cdots, x_n)$，则有

$$a_{ii}\boldsymbol{x}_i>\sum_{j\neq i} |a_{ij}| x_j, \quad i=1, \cdots, n$$

这表明 $\boldsymbol{AD}$ 是严格行对角占优矩阵，且 $\boldsymbol{AD}$ 的所有对角线元素为正数.

(iii)⇒(iv) 设 $\boldsymbol{D}=\mathrm{diag}(d_1, \cdots, d_n)>0$ 使 $\boldsymbol{AD}$ 为严格行对角占优矩阵，且 $\boldsymbol{AD}$ 的所有对角线元素为正数. 于是

$$a_{ii}>\frac{1}{d_i}\sum_{j\neq i} |a_{ij}| d_j, \quad i=1, \cdots, n$$

由此，据圆盘定理(参见 7.1 节)知 $\boldsymbol{A}$ 的每个实特征值均为正数.

(iv)⇒(v) 设 $\boldsymbol{A}=s\boldsymbol{E}-\boldsymbol{B}$,其中 $s>0$, $\boldsymbol{B}\geqslant 0$,则 $s-\rho(\boldsymbol{B})$ 为矩阵 $\boldsymbol{A}$ 的特征值，所以 $s>\rho(\boldsymbol{B})$，故 $\boldsymbol{A}$ 为非奇异 M 矩阵.

(v)⇒(vi) 由定理 7.7.1 即得.

(vi)⇒(vii) 设 $\boldsymbol{A}_k$ 为 $\boldsymbol{A}$ 的任一 $k$ 阶主子阵，$\lambda$ 为 $\boldsymbol{A}_k$ 的实特征值. 下面证明 $\lambda$ 为正数. 假设 $\lambda\leqslant 0$,定义矩阵 $\boldsymbol{B}=(b_{ij})$ 如下：

$$b_{ij}=\begin{cases}a_{ii}-\lambda, & i=j\\ a_{ij}, & i\neq j \text{ 且 } i, j \text{ 为 } \boldsymbol{A}_k \text{ 在 } \boldsymbol{A} \text{ 中的行,列序数}\\ 0, & \text{其他情形}\end{cases}$$

于是 $\boldsymbol{B}\geqslant \boldsymbol{A}$. 由(vi)知 $\boldsymbol{B}$ 非奇异. 因为 $\lambda$ 为 $\boldsymbol{A}_k$ 的实特征值，所以

$$|\boldsymbol{B}_k|=|\boldsymbol{A}_k-\lambda\boldsymbol{E}_k|$$

这里 $\boldsymbol{B}_k$ 为 $\boldsymbol{B}$ 的主子阵，其在 $\boldsymbol{B}$ 中的行、列序数与 $\boldsymbol{A}_k$ 相同. 于是

$$|\boldsymbol{B}|=\prod_{i\notin S}b_{ii}\,|\boldsymbol{B}_k|=0$$

其中 $S$ 表示 $\boldsymbol{B}_k$ 在 $\boldsymbol{B}$ 的行数的集合. 由 $|\boldsymbol{B}|=0$ 知 $\boldsymbol{B}$ 是奇异的，这与 $\boldsymbol{B}$ 非奇异矛盾.

(vii)⇒(viii) 设 $\boldsymbol{A}_k$ 为 $\boldsymbol{A}$ 的任一 $k$ 阶主子阵，$|\boldsymbol{A}_k|=\lambda_1\cdots\lambda_k$，其中 $\lambda_i$, $i=1,\cdots,k$ 为特征多项式 $f_k(\lambda)=|\lambda\boldsymbol{E}_k-\boldsymbol{A}_k|$ 的根. 由于实系数多项式的虚根成对出现，所以 $|\boldsymbol{A}_k|>0$.

(viii)⇒(ix) 显然成立.

(ix)⇒(iv) $|\boldsymbol{A}-\lambda\boldsymbol{E}|=(-\lambda)^n+c_1(-\lambda)^{n-1}+\cdots+c_n$，其中 $c_k$ 为 $\boldsymbol{A}$ 的所有 $k$ 阶主子式之和. 由(ix)知 $c_k>0$，所以多项式 $|\boldsymbol{A}-\lambda\boldsymbol{E}|$ 不能有非正的实根，亦即 $\boldsymbol{A}$ 的所有实特征值均为整数.

(v)⇒(x) 设 $\boldsymbol{A}=s\boldsymbol{E}-\boldsymbol{B}$，其中 $s>\rho(\boldsymbol{B})$，$\boldsymbol{B}\geqslant 0$. 令 $\boldsymbol{P}=s\boldsymbol{E}$，$\boldsymbol{Q}=\boldsymbol{B}$，则(x)成立.

(x)⇒(xi) 设 $\boldsymbol{A}=\boldsymbol{P}-\boldsymbol{Q}$ 为(x)中的分裂，则

$$\boldsymbol{A}=\boldsymbol{P}(\boldsymbol{E}-\boldsymbol{C})$$

其中

$$\boldsymbol{C}=\boldsymbol{P}^{-1}\boldsymbol{Q}$$

因为 $\rho(\boldsymbol{C})<1$，所以

$$\boldsymbol{A}^{-1}=(\boldsymbol{E}-\boldsymbol{C})^{-1}\boldsymbol{P}^{-1}=(\boldsymbol{E}+\boldsymbol{C}+\boldsymbol{C}^2+\cdots)\boldsymbol{P}^{-1}$$

从而由 $\boldsymbol{C}=\boldsymbol{P}^{-1}\boldsymbol{Q}\geqslant 0$ 知

$$A^{-1} \geqslant 0$$

(xi)⇒(i) 令 $x = A^{-1}x_0, x_0 = (1, \cdots, 1)^T$. 由 $A^{-1} \geqslant 0$ 知 $x \geqslant 0$,满足 $Ax > 0$.

以上讨论了非奇异的M矩阵的一些性质,一般的M矩阵在应用中一样重要,下面给出一般的M矩阵的一些特性.

**定理 7.7.3** 设 $A \in Z^{n\times n}$,则下列各命题彼此等价:

(i) $A$ 为M矩阵;

(ii) 对每个 $\varepsilon > 0$, $A + \varepsilon E$ 为非奇异M矩阵;

(iii) $A$ 的每个主子阵的所有实特征值非负;

(iv) $A$ 的所有主子式非负;

(v) 对每个 $k(1 \leqslant k \leqslant n)$, $A$ 的所有 $k$ 阶主子式之和为非负实数;

(vi) $A$ 的所有实特征值非负.

证明略.

最后,对于 $A$ 为不可约的奇异M矩阵,有如下的结果(证略).

**定理 7.7.4** 设 $A$ 为不可约奇异M矩阵,则

(i) $r(A) = n - 1$;

(ii) 存在正向量 $x$ 使 $Ax = 0$;

(iii) $A$ 的所有真主子阵为非奇异M矩阵,特别地,有

$$a_{ii} > 0, \quad 1 \leqslant i \leqslant n$$

(iv) 若 $x \in \mathbf{R}^n$ 满足 $Ax \geqslant 0$,则

$$Ax = 0$$

## 习题 7

1. 试举两个例子,说明两个相交的盖尔圆构成的连通区域可以在每个盖尔圆中各有一个特征值,或只在一个盖尔圆中有两个特征值而另一个不含特征值.

2. 试利用圆盘定理确定矩阵

$$A = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 5 & 1 \\ -2 & -1 & 9 \end{pmatrix}$$

特征值的分布范围. 对盖尔圆进行适当的缩放,判断 $A$ 的特征值是实数还是复数.

3. 若 $\lambda_1, \lambda_2, \cdots, \lambda_n$ 是矩阵 $A$ 的特征值, $\mu_1, \mu_2, \cdots, \mu_n$ 是 $A^HA$ 的特征值.

证明：

$$\sum_{i=1}^{n} |\lambda_i|^2 \leqslant \sum_{i=1}^{n} \mu_i$$

4. 设 $f(z) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n$，$a_0 \neq 0$. 证明：若 $z_0$ 是 $f$ 的任一零点，则有

$$|z_0| \leqslant 1 + \max_{1 \leqslant k \leqslant n} \left| \frac{a_k}{a_0} \right|$$

5. 设 $\boldsymbol{H} = (h_{ij})_{n\times n} \in \mathbf{C}^{n\times n}$ 是 Hermite 矩阵，$\lambda_1$，$\lambda_n$ 分别是 $\boldsymbol{H}$ 的最小值、最大值. 证明：

(1) $\lambda_1 \leqslant h_{jj} \leqslant \lambda_n$，$j = 1, 2, \cdots, n$；

(2) $\lambda_1 \leqslant n^{-1} \sum_{i,j=1}^{n} h_{ij} \leqslant \lambda_n$.

6. 试证明式(7.3.4).

7. 试证明式(7.3.6).

8. 设 $n$ 阶 Hermite 矩阵 $\boldsymbol{A}_n$ 的特征值是 $\lambda_1 \leqslant \lambda_2 \leqslant \cdots \leqslant \lambda_n$，$\boldsymbol{A}_n$ 的最后一行和最后一列去掉所得的子矩阵 $\boldsymbol{A}_{n-1}$ 的最大特征值是 $\mu$. 证明：$\lambda_1 \leqslant \mu \leqslant \lambda_2$.

9. 设 $\boldsymbol{A}$ 为 $n$ 阶非负矩阵，且对某个 $k$，$\boldsymbol{A}^k > 0$，证明：$\rho(\boldsymbol{A}) > 0$.

10. 设 $\boldsymbol{A}$ 为非零的非负矩阵，若 $\boldsymbol{A}$ 有正的特征向量，证明：$\rho(\boldsymbol{A}) > 0$.

11. 设 $\boldsymbol{A}$ 为 $n$ 阶非负矩阵，证明：若 $\boldsymbol{A}$ 的各个行和是常数，则 $\rho(\boldsymbol{A}) = \|\boldsymbol{A}\|_\infty$；若 $\boldsymbol{A}$ 的各个列和是常数，则 $\rho(\boldsymbol{A}) = \|\boldsymbol{A}\|_1$.

12. 设 $\boldsymbol{A}$ 为 $n$ 阶非负矩阵且有正向量 $\boldsymbol{y}$ 使得 $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} = \rho(\boldsymbol{A})\boldsymbol{y}$. 证明：若存在 $\boldsymbol{x} \geqslant 0$，$\boldsymbol{x} \neq \boldsymbol{0}$ 使得 $\boldsymbol{A}\boldsymbol{x} \geqslant \rho(\boldsymbol{A})\boldsymbol{x}$，则 $\boldsymbol{A}\boldsymbol{x} = \rho(\boldsymbol{A})\boldsymbol{x}$.

# 习题答案与提示

## 习　题　1

1. 都是线性空间.
2. (2),(3)是子空间;(1)不是.
3. (1) 不是;(2) 是.
4. (1) 由子空间定义可证;

   (2) $W$ 的基为 $\begin{pmatrix}1 & 0\\0 & -1\end{pmatrix}$, $\begin{pmatrix}0 & 1\\0 & 0\end{pmatrix}$, $\begin{pmatrix}0 & 0\\1 & 0\end{pmatrix}$;

   (3) $(3, 2, 5)^{\mathrm{T}}$.
5. (1) 由子空间的判定定理可证;

   (2) $\begin{pmatrix}1 & 0\\0 & 1\end{pmatrix}$, $\begin{pmatrix}0 & 1\\0 & 0\end{pmatrix}$为 $C(\boldsymbol{A})$的基, $\dim C(\boldsymbol{A}) = 2$;
6. (1) $\boldsymbol{P} = \begin{pmatrix}-1 & 0 & 2 & 1\\0 & 3 & 1 & -3\\0 & -1 & 0 & 0\\2 & 4 & 1 & 2\end{pmatrix}$;

   (2) $\boldsymbol{P}^{-1}$;

   (3) $\boldsymbol{A}$ 在$\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_3, \boldsymbol{\varepsilon}_4$ 下的坐标为$(-1, 3, 0, 2)^{\mathrm{T}}$, $\boldsymbol{A}$ 在$\boldsymbol{\eta}_1, \boldsymbol{\eta}_2, \boldsymbol{\eta}_3, \boldsymbol{\eta}_4$ 下的坐标为$\left(\frac{28}{19}, 0, \frac{12}{19}, -\frac{15}{19}\right)^{\mathrm{T}}$.
7. 因为

$$(2, -1, 3, 3) = (-1)(1, 1, 0, 0) + 3(1, 0, 1, 1)$$

$$(0, 1, -1, -1) = (1, 1, 0, 0) + (-1)(1, 0, 1, 1)^{\mathrm{T}}$$

8. 设 $\boldsymbol{\alpha} = (a_1, a_2, a_3, a_4) \in W_1 \cap W_2$, 则有

$$a_1 - a_2 + a_3 - a_4 = 0,\ a_1 + a_2 + a_3 + a_4 = 0$$

   可求得

$$\eta_1 = (1,\ 0,\ -1,\ 0), \qquad \eta_2 = (0.1,\ 0,\ -1)$$

为 $W_1 \cap W_2$ 的基.

9. 由定义可证 $T_1$，$T_2$都是线性变换，且

$$\begin{aligned}(T_1 + T_2)(x_1,\ x_2) &= T_1(x_1,\ x_2) + T_2(x_1,\ x_2) \\ &= (x_2,\ -x_1) + (x_1,\ -x_2) \\ &= (x_1 + x_2,\ -x_1 - x_2)\end{aligned}$$

10. (1) 由定义可证；

(2) $\begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$.

11. $T$ 在基 $E_{11} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, $E_{12} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $E_{21} = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $E_{22} = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$ 下的矩阵为

$$A = \begin{pmatrix} 0 & 4 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 4 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$A$ 的特征值与特征向量分别为

$$\begin{aligned} &\lambda_1 = \lambda_2 = 2, && \lambda_3 = \lambda_4 = -2 \\ &p_1 = (2,\ 1,\ 0,\ 0)^T, && p_2 = (0,\ 0,\ 2,\ 1)^T \\ &p_3 = (2,\ -1,\ 0,\ 0)^T, && p_4 = (0,\ 0,\ 2,\ -1)^T \end{aligned}$$

令 $P = (p_1,\ p_2,\ p_3,\ p_4)$，则

$$P^{-1}AP = \Lambda = \mathrm{diag}\{2,\ 2,\ -2,\ -2\}$$

令

$$\beta_i = (E_{11},\ E_{12},\ E_{21},\ E_{22})p_i, \quad i = 1,\ 2,\ 3,\ 4$$

则 $\beta_1$, $\beta_2$, $\beta_3$, $\beta_4$ 为 $\mathbf{R}^{2\times 2}$ 的基，且 $T$ 在此基下的矩阵为 $\Lambda$.

12. $\begin{pmatrix} 0 & 1 & -1 \\ 1 & 1 & 1 \\ 3 & 0 & 2 \end{pmatrix}$.

13. (1) 由定义可证；

(2) $\begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$；

(3) $T$ 在基

$$\boldsymbol{\alpha}_1 = 1+t+t^2, \quad \boldsymbol{\alpha}_2 = 1-t, \quad \boldsymbol{\alpha}_3 = 1-t^2$$

下矩阵为对角矩阵.

14. $\begin{bmatrix} 0 & -1 & 1 \\ -1 & -1 & 2 \\ -1 & -1 & 3 \end{bmatrix}$.

## 习　题　2

1. (1) 不是； (2) 不是； (3) 是； (4) 不是.
2. 验证内积定义中的 4 个条件都成立即可.
3. 充分性由内积定义可证.

必要性：

(1) 证明 $\boldsymbol{A}$ 是对称阵，即证明 $a_{ij} = a_{ji}$.

取 $\boldsymbol{\alpha} = \boldsymbol{\varepsilon}_i = (0, \cdots, 1, 0, \cdots, 0)^{\mathrm{T}}$，$\boldsymbol{\beta} = \boldsymbol{\varepsilon}_j = (0, \cdots, 1, \cdots, 0)^{\mathrm{T}}$，由内积定义有

$$(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = \boldsymbol{\varepsilon}_i^{\mathrm{T}} \boldsymbol{A} \boldsymbol{\varepsilon}_j = \boldsymbol{a}_{ij}$$

$$(\boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}_i) = \boldsymbol{\varepsilon}_j^{\mathrm{T}} \boldsymbol{A} \boldsymbol{\varepsilon}_i = \boldsymbol{a}_{ji}$$

又由内积定义的条件(i)知

$$(\boldsymbol{\varepsilon}_i, \boldsymbol{\varepsilon}_j) = (\boldsymbol{\varepsilon}_j, \boldsymbol{\varepsilon}_i)$$

所以

$$a_{ij} = a_{ji}$$

故 $\boldsymbol{A}$ 是对称矩阵.

(2) 证明 $\boldsymbol{A}$ 正定.

由内积定义条件(iv)知，对任意 $\boldsymbol{\alpha} \in \mathbf{R}^n$，$\boldsymbol{\alpha} \neq 0$，有

$$(\boldsymbol{\alpha}, \boldsymbol{\alpha}) = \boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{\alpha} > 0$$

从而可知 $\boldsymbol{\alpha}^{\mathrm{T}} \boldsymbol{A} \boldsymbol{\alpha}$ 是正定二次型，故 $\boldsymbol{A}$ 为正定矩阵.

4. (1) 根据条件，$\boldsymbol{\gamma}$ 可以表示为 $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \cdots, \boldsymbol{\alpha}_n$ 的线性组合，即

$$\boldsymbol{\gamma}=k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_n\boldsymbol{\alpha}_n$$

于是

$$\begin{aligned}(\boldsymbol{\gamma},\ \boldsymbol{\gamma})&=(\boldsymbol{\gamma},\ k_1\boldsymbol{\alpha}_1+k_2\boldsymbol{\alpha}_2+\cdots+k_n\boldsymbol{\alpha}_n)\\&=k_1(\boldsymbol{\gamma},\ \boldsymbol{\alpha}_1)+k_2(\boldsymbol{\gamma},\ \boldsymbol{\alpha}_2)+\cdots+k_n(\boldsymbol{\gamma},\ \boldsymbol{\alpha}_n)=0\end{aligned}$$

故

$$\boldsymbol{\gamma}=\mathbf{0}$$

(2) 由题设知

$$(\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_2,\ \boldsymbol{\alpha}_i)=(\boldsymbol{\gamma}_1,\ \boldsymbol{\alpha}_i)-(\boldsymbol{\gamma}_2,\ \boldsymbol{\alpha}_i),\quad i=1,\ 2,\ \cdots,\ n$$

由(1)的结果知

$$\boldsymbol{\gamma}_1-\boldsymbol{\gamma}_2=\mathbf{0}$$

即

$$\boldsymbol{\gamma}_1=\boldsymbol{\gamma}_2$$

5. 解空间的标准正交基为

$$\boldsymbol{\gamma}_1=\frac{1}{\sqrt{2}}(0,\ 1,\ 1,\ 0,\ 0)^{\mathrm{T}}$$

$$\boldsymbol{\gamma}_2=\frac{1}{\sqrt{10}}(-2,\ 1,\ -1,\ 2,\ 0)^{\mathrm{T}}$$

$$\boldsymbol{\gamma}_3=\frac{1}{3\sqrt{35}}(7,\ -6,\ 6,\ 13,\ 5)^{\mathrm{T}}$$

6. (1) 显然 $0\in V_1$，故 $V_1$ 非空. 对任意 $\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2\in V_1$ 和 $k\in\mathbf{R}$，有

$$(\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2,\ \boldsymbol{\alpha})=(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha})+(\boldsymbol{\alpha}_2,\ \boldsymbol{\alpha})=0$$

$$(k\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha})=k(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha})=0$$

即

$$\boldsymbol{\alpha}_1+\boldsymbol{\alpha}_2\in V_1,\quad k\boldsymbol{\alpha}_1\in V_1$$

因此 $V_1$ 是 $V$ 的子空间.

(2) 由于 $\boldsymbol{\alpha}\neq\mathbf{0}$ 是线性无关的，因此可以将 $\boldsymbol{\alpha}$ 扩充为 $V$ 的一组正交基 $\boldsymbol{\alpha}_2$，$\boldsymbol{\alpha}_3,\ \cdots,\ \boldsymbol{\alpha}_n$，因为

$$(\boldsymbol{\alpha}_i,\ \boldsymbol{\alpha})=0,\quad i=2,\ 3,\ \cdots,\ n$$

所以

$$\boldsymbol{\alpha}_i \in V_1, \quad i = 2, 3, \cdots, n$$

对 $\forall \boldsymbol{\beta} \in V_1$，有 $\boldsymbol{\beta} \in V$，于是

$$\boldsymbol{\beta} = k_1 \boldsymbol{\alpha} + k_2 \boldsymbol{\alpha}_2 + \cdots + k_n \boldsymbol{\alpha}_n$$

而

$$0 = (\boldsymbol{\beta}, \boldsymbol{\alpha}) = k_1(\boldsymbol{\alpha}, \boldsymbol{\alpha}) + k_2(\boldsymbol{\alpha}_2, \boldsymbol{\alpha}) + \cdots + k_n(\boldsymbol{\alpha}_n, \boldsymbol{\alpha}) = k_1(\boldsymbol{\alpha}, \boldsymbol{\alpha})$$

故

$$k_1 = 0$$

即 $\boldsymbol{\beta}$ 可以由 $\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3 \cdots, \boldsymbol{\alpha}_n$ 线性表示. 从而 $\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3, \cdots, \boldsymbol{\alpha}_n$ 是 $V_1$ 的基.

$$\dim V_1 = n - 1$$

7. 因 $W^{\perp} = \{\boldsymbol{x} \in \mathbf{R}^3 \mid (\boldsymbol{x}, \boldsymbol{\alpha}) = 0, \boldsymbol{x} = (x_1, x_2, x_3)\}$，故解方程

$$x_1 - 2x_2 + x_3 = 0$$

得

$$\boldsymbol{x} = (2a - b, a, b)^{\mathrm{T}}, \quad a, b \in \mathbf{R}$$

即

$$W^{\perp} = \{\boldsymbol{x} \in \mathbf{R}^3 \mid \boldsymbol{x} = (2a - b, a, b)^{\mathrm{T}}, a, b \in \mathbf{R}\}$$

8. (1) 由 $\boldsymbol{x} \in (V_0^{\perp})^{\perp}$ 得 $\boldsymbol{x} \perp V_0^{\perp}$，故对所有的 $\boldsymbol{y} \in V_0^{\perp}$，$(\boldsymbol{x}, \boldsymbol{y}) = 0$，故 $\boldsymbol{x} \in V_0$.

(2) 任取 $\boldsymbol{x} \in V_2^{\perp}$，则对所有的 $\boldsymbol{y} \in V_2$，有 $(\boldsymbol{x}, \boldsymbol{y}) = 0$. 但 $V_1 \subset V_2$，故当 $\boldsymbol{y} \in V_1$ 时，必然有 $(\boldsymbol{x}, \boldsymbol{y}) = 0$，所以，$x \in V_1^{\perp}$.

(3) 由于 $V_1, V_2 \subset V_1 + V_2$，由(2)可知 $(V_1 + V_2)^{\perp} \subset V_1^{\perp} \cap V_2^{\perp}$；反之，任取 $\boldsymbol{y} \in V_1^{\perp} \cap V_2^{\perp}$，则当 $\boldsymbol{x}_1 + \boldsymbol{x}_2 \in V_1 + V_2$ 时，有

$$(\boldsymbol{x}_1 + \boldsymbol{x}_2, \boldsymbol{y}) = (\boldsymbol{x}_1, \boldsymbol{y}) + (\boldsymbol{x}_2, \boldsymbol{y}) = 0 + 0 = 0$$

这说明 $\boldsymbol{y} \in (V_1 + V_2)^{\perp}$. 所以又有

$$V_1^{\perp} \cap V_2^{\perp} \subset (V_1 + V_2)^{\perp}$$

(4) 利用上一小题已得的结果可知

$$(V_1^{\perp} + V_2^{\perp})^{\perp} = V_1^{\perp\perp} \cap V_2^{\perp\perp} = V_1 \cap V_2$$

在上式两边分别取正交补，即有

$$V_1^{\perp} + V_2^{\perp} = (V_1 \cap V_2)^{\perp}$$

9. 设方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 有解 $\boldsymbol{x}_0$，则对齐次线性方程组 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{y}=\boldsymbol{0}$ 的任一解向量 $\boldsymbol{y}_0$，有

$$(\boldsymbol{y}_0,\ \boldsymbol{b})=(\boldsymbol{y}_0,\ \boldsymbol{Ax}_0)=(\boldsymbol{A}^{\mathrm{H}}\boldsymbol{y}_0,\ \boldsymbol{x}_0)=(\boldsymbol{0},\ \boldsymbol{x}_0)=0$$

上式表明，向量 $\boldsymbol{b}$ 与齐次线性方程组的每一个解向量正交；反之，若设齐次线性方程组的解空间为 $\boldsymbol{W}$，矩阵 $\boldsymbol{A}$ 的列向量组是 $\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n$，则对任意的 $\boldsymbol{y}\in\boldsymbol{W}$，必有

$$(\boldsymbol{y},\ \boldsymbol{\alpha}_i)=0,\quad i=1,\ 2,\ \cdots,\ n$$

即

$$\boldsymbol{y}\in[\mathrm{Span}(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n)]^{\perp}$$

同理，对每个

$$\boldsymbol{y}\in[\mathrm{Span}(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n)]^{\perp}$$

由 $(\boldsymbol{y},\ \boldsymbol{\alpha}_i)=0(i=1,\ 2,\ \cdots,\ n)$ 可知，$\boldsymbol{y}\in\boldsymbol{W}$. 从而

$$\boldsymbol{W}=[\mathrm{Span}(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n)]^{\perp}$$

因此，当设 $\boldsymbol{b}\perp\boldsymbol{W}$ 时，由

$$\boldsymbol{b}\in\boldsymbol{W}^{\perp}=[\mathrm{Span}(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n)]^{\perp\perp}=[\mathrm{Span}(\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n)]$$

可知，$\boldsymbol{b}$ 可以由向量组 $\boldsymbol{\alpha}_1,\ \boldsymbol{\alpha}_2,\ \cdots,\ \boldsymbol{\alpha}_n$ 线性表示. 所以方程组 $\boldsymbol{Ax}=\boldsymbol{b}$ 有解.

10. $\forall\boldsymbol{\alpha}\in V_1$，则 $T\boldsymbol{\alpha}=\boldsymbol{\alpha}$；$\forall\boldsymbol{\beta}\in V_2$，则

$$\boldsymbol{\beta}=\boldsymbol{x}-T\boldsymbol{x}$$

因为 $T$ 为正交变换，故

$$(T\boldsymbol{\alpha},\ T\boldsymbol{x})=(\boldsymbol{\alpha},\ \boldsymbol{x})$$

于是

$$\begin{aligned}(\boldsymbol{\alpha},\ \boldsymbol{\beta})&=(T\boldsymbol{\alpha},\ \boldsymbol{x}-T\boldsymbol{x})=(T\boldsymbol{\alpha},\ \boldsymbol{x})-(T\boldsymbol{\alpha},\ T\boldsymbol{x})\\&=(\boldsymbol{\alpha},\ \boldsymbol{x})-(\boldsymbol{\alpha},\ \boldsymbol{x})=\boldsymbol{0}\end{aligned}$$

故
$$\boldsymbol{\alpha}\in V_2^{\perp}$$

另一方面，$\forall\boldsymbol{x}\in V_2^{\perp}$，则

$$(\boldsymbol{x},\ \boldsymbol{x}-T\boldsymbol{x})=(\boldsymbol{x},\ \boldsymbol{x})-(\boldsymbol{x},\ T\boldsymbol{x})=0$$

由此

$$(\boldsymbol{x}-T\boldsymbol{x},\ \boldsymbol{x}-T\boldsymbol{x})=(\boldsymbol{x},\ \boldsymbol{x})+(T\boldsymbol{x},\ T\boldsymbol{x})-2(\boldsymbol{x},\ T\boldsymbol{x})=0$$

故

$$\| x - Tx \| = 0$$

从而

$$x - Tx = \mathbf{0}$$

即

$$x \in V_1$$

11. 易证(略).

12. 显然 $A$ 满足 $A^H = A$，即 $A$ 是 Hermite 矩阵，从而 $A$ 是正规矩阵. 又

$$|\lambda E - A| = (\lambda - 1)(\lambda + 1)(\lambda + 2)$$

所以 $A$ 的特征值为

$$\lambda_1 = 1, \quad \lambda_2 = -1, \quad \lambda_3 = -2$$

可以求得对应的特征向量分别为

$$\eta_1 = (1, -2i, 1)^T, \quad \eta_2 = (-1, 0, 1)^T, \quad \eta_3 = (1, -i, 1)^T$$

它们是两两正交的，单位化得

$$\gamma_1 = \left(\frac{1}{\sqrt{6}}, -\frac{2}{\sqrt{6}}i, \frac{1}{\sqrt{6}}\right)^T, \quad \gamma_2 = \left(-\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}}\right)^T$$

$$\gamma_3 = \left(\frac{1}{\sqrt{3}}, \frac{-i}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)^T$$

于是酉矩阵

$$U = \begin{pmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ -\frac{2i}{\sqrt{6}} & 0 & \frac{-i}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix}$$

使得

$$U^{-1}AU = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -2 \end{pmatrix}$$

13. 由

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(x_1-x_2)=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}=\boldsymbol{0}$$

知

$$(x_1-x_2)^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}(x_1-x_2)=0$$

于是

$$\boldsymbol{A}(x_1-x_2)=\boldsymbol{0}$$

## 习　题　3

1. (1) 行列式因子

$$D_1(\lambda)=D_2(\lambda)=D_3(\lambda)=1$$
$$D_4(\lambda)=(\lambda+1)(\lambda+2)(\lambda-1)(\lambda-2)$$

故不变因子

$$d_1(\lambda)=d_2(\lambda)=d_3(\lambda)$$
$$d_4(\lambda)=(\lambda+1)(\lambda+2)(\lambda-1)(\lambda-2)$$

所以初级因子为

$$\lambda+1,\ \lambda+2,\ \lambda-1,\ \lambda-2$$

(2)
$$d_1(\lambda)=d_2(\lambda)=\cdots=d_{n-1}(\lambda)=1$$
$$d_n(\lambda)=(\lambda-a)^n$$

所以初级因子为

$$(\lambda-a)^n$$

(3) 不变因子

$$d_1(\lambda)=d_2(\lambda)=1,\ d_3(\lambda)=(\lambda-2)(\lambda-1)^2$$

所以初级因子为

$$\lambda-2,\ (\lambda-1)^2$$

2. (1) $\begin{bmatrix}1 & & \\ & 1 & \\ & & (\lambda-2)^2\end{bmatrix}$;

(2) $\begin{bmatrix}1 & 0 & 0\\ 0 & \lambda(\lambda+1) & 0\\ 0 & 0 & \lambda(\lambda+1)^2\end{bmatrix}$.

3. $$f(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}|=\lambda^n+a_1\lambda^{n-1}+a_2\lambda^{n-2}+\cdots+(-1)^n|\boldsymbol{A}|$$

由 $f(\boldsymbol{A})=0$,得

$$\boldsymbol{A}^n+a_1\boldsymbol{A}^{n-1}+a_2\boldsymbol{A}^{n-2}+\cdots+(-1)^n|\boldsymbol{A}|\boldsymbol{E}=\boldsymbol{0}$$

又因为 $\boldsymbol{A}$ 可逆,$|\boldsymbol{A}|\neq 0$,故有

$$\boldsymbol{A}^{-1}=\frac{1}{(-1)^{n+1}|\boldsymbol{A}|}(\boldsymbol{A}^{n-1}+a_1\boldsymbol{A}^{n-2}+\cdots+a_{n-1}\boldsymbol{E})$$

4. $$f(\lambda)=|\lambda\boldsymbol{E}-\boldsymbol{A}|=\lambda^2-6\lambda+7$$

又作多项式

$$\varphi(\lambda)=2\lambda^4-12\lambda^3+19\lambda^2-29\lambda+37$$

则

$$\varphi(\lambda)=f(\lambda)(2\lambda^2+5)+\lambda+2$$

由 $f(\boldsymbol{A})=0$,得

$$\varphi(\boldsymbol{A})=\boldsymbol{A}+2\boldsymbol{E}$$

即

$$\boldsymbol{B}=2\boldsymbol{A}^4-12\boldsymbol{A}^3+19\boldsymbol{A}^2-29\boldsymbol{A}+37\boldsymbol{E}$$
$$=\varphi(\boldsymbol{A})=\boldsymbol{A}+2\boldsymbol{E}=\begin{pmatrix}3 & -1\\ 2 & 7\end{pmatrix}$$

且 $\begin{vmatrix}3 & -1\\ 2 & 7\end{vmatrix}\neq 0$,所以 $\boldsymbol{B}$ 可逆. 又

$$|\lambda\boldsymbol{E}-\boldsymbol{B}|=\lambda^2-10\lambda+23$$
$$\boldsymbol{B}^2-10\boldsymbol{B}+23\boldsymbol{E}=\boldsymbol{0}$$

即

$$(\boldsymbol{A}+2\boldsymbol{E})^2-10(\boldsymbol{A}+2\boldsymbol{E})+23\boldsymbol{E}=\boldsymbol{0}$$
$$(\boldsymbol{A}+2\boldsymbol{E})(\boldsymbol{A}-8\boldsymbol{E})=-23\boldsymbol{E}$$

故所求逆矩阵为

$$\boldsymbol{B}^{-1}=(\boldsymbol{A}+2\boldsymbol{E})^{-1}=-\frac{1}{23}(\boldsymbol{A}-8\boldsymbol{E})$$

5. $f(\boldsymbol{A})=\begin{pmatrix}-3 & 48 & -26\\ 0 & 95 & -61\\ 0 & -61 & 34\end{pmatrix}$.

6. (1) $\begin{pmatrix}2&0&0\\0&0&0\\0&0&0\end{pmatrix}$； (2) $\begin{pmatrix}1&0&0\\0&\mathrm{i}&0\\0&0&-\mathrm{i}\end{pmatrix}$； (3) $\begin{pmatrix}-1&0&0\\0&-1&1\\0&0&-1\end{pmatrix}$；

(4) $\begin{pmatrix}1&1&0\\0&1&1\\0&0&1\end{pmatrix}$； (5) $\begin{pmatrix}1&1&0&0\\0&1&0&0\\0&0&1&1\\0&0&0&1\end{pmatrix}$.

7. 只需证 Jordan 块，$\boldsymbol{J}_i$ 与 $\boldsymbol{J}_i^{\mathrm{T}}$ 相似.

取 $\boldsymbol{P}_i=\begin{pmatrix}0&&&&1\\&&&1&\\&&\iddots&&\\&1&&&\\1&&&&0\end{pmatrix}$，验证

$$\boldsymbol{P}_i^{-1}\boldsymbol{J}_i\boldsymbol{P}_i=\boldsymbol{J}_i^{\mathrm{T}}$$

8.
$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}=\boldsymbol{J},\quad \boldsymbol{A}=\boldsymbol{P}\boldsymbol{J}\boldsymbol{P}^{-1},\quad \boldsymbol{A}^k=\boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}$$

其中

$$\boldsymbol{P}=\begin{pmatrix}1&0&0\\-1&-1&1\\2&1&0\end{pmatrix},\qquad \boldsymbol{J}=\begin{pmatrix}1&1&0\\0&1&0\\0&0&2\end{pmatrix}$$

$$\boldsymbol{P}^{-1}\begin{pmatrix}1&0&0\\-2&0&1\\-1&1&1\end{pmatrix},\qquad \boldsymbol{J}^k=\begin{pmatrix}1&k&0\\0&1&0\\0&0&2^k\end{pmatrix}$$

故

$$\boldsymbol{A}^k=\boldsymbol{P}\boldsymbol{J}^k\boldsymbol{P}^{-1}=\begin{pmatrix}-2k+1&0&k\\2k+1-2^k&2^k&-k-1+2^k\\-4k&0&2k+1\end{pmatrix}$$

9. $\begin{cases}x_1=[c_1(t+1)+2c_2t-c_3t]\mathrm{e}^t\\x_2=[-c_1t+c_2(1-2t)+c_3t]\mathrm{e}^t\\x_3=[-c_1t-2c_2t+c_3(1+t)]\mathrm{e}^t\end{cases}$ $(C_1,C_2,C_3\in\mathbf{C})$.

10. $m_A(x)=(\lambda-1)^2$.

11. 由题 $(\boldsymbol{A}-2\boldsymbol{E})(\boldsymbol{A}-3\boldsymbol{E})=\boldsymbol{0}$, 则

$$\varphi(\lambda)=(\lambda-2)(\lambda-3)$$

是 $\boldsymbol{A}$ 的零化多项式. 又 $\boldsymbol{A}$ 的最小多项式 $m(\lambda)$ 整除任一零化多项式, 故 $m(\lambda)$ 无重根, 从而 $\boldsymbol{A}$ 可以对角化.

12. (1)

$$m_A(\lambda)=(\lambda-2)(\lambda-3)$$

$$m_A(\lambda)=(\lambda-2)^2(\lambda-3)$$

$$m_A(\lambda)=(\lambda-2)(\lambda-3)^2$$

$$m_A(\lambda)=(\lambda-2)^2(\lambda-3)^2$$

(2)

$$\begin{pmatrix}2&&&\\&2&&\\&&3&\\&&&3\end{pmatrix},\quad\begin{pmatrix}2&1&&\\&2&&\\&&3&\\&&&3\end{pmatrix}$$

$$\begin{pmatrix}2&&&\\&2&&\\&&3&1\\&&&3\end{pmatrix},\quad\begin{pmatrix}2&1&&\\&2&&\\&&3&1\\&&&3\end{pmatrix}$$

## 习　题　4

1. $\boldsymbol{A}=\begin{pmatrix}1&0&0\\2&1&0\\-1&-3&1\end{pmatrix}\begin{pmatrix}2&1&1\\0&-1&-2\\0&0&-4\end{pmatrix}$.

2. (1) $\begin{pmatrix}1&0\\0&1\end{pmatrix}\begin{pmatrix}1&1\\0&1\end{pmatrix}$;

(2) $\begin{pmatrix}\frac{1}{\sqrt{2}}&-\frac{1}{\sqrt{6}}\\\frac{1}{\sqrt{2}}&\frac{1}{\sqrt{6}}\\0&\frac{2}{\sqrt{6}}\end{pmatrix}\begin{pmatrix}\sqrt{2}&\frac{1}{\sqrt{2}}\\0&\frac{\sqrt{3}}{\sqrt{2}}\end{pmatrix}$;

(3) $\begin{pmatrix} 0 & \frac{4}{5} & -\frac{3}{5} \\ 1 & 0 & 0 \\ 0 & -\frac{3}{5} & \frac{4}{5} \end{pmatrix}\begin{pmatrix} 1 & 1 & 1 \\ 0 & 5 & 2 \\ 0 & 0 & 1 \end{pmatrix}$.

3. $\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 0 & 2 \\ -2 & 4 \end{pmatrix}\begin{pmatrix} 1 & 0 & 2 & 1 \\ 0 & 1 & \frac{1}{2} & -\frac{1}{2} \end{pmatrix}$.

4. (1) $\boldsymbol{A}$ 是正规矩阵的充要条件是 $\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = \boldsymbol{A}\boldsymbol{A}^{\mathrm{H}}$，则

$$\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A} = (\boldsymbol{B}^{\mathrm{T}} - \mathrm{i}\boldsymbol{C}^{\mathrm{T}})(\boldsymbol{B} + \mathrm{i}\boldsymbol{C}) = \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} + \mathrm{i}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{C} - \boldsymbol{C}^{\mathrm{T}}\boldsymbol{B})$$

$$\boldsymbol{A}\boldsymbol{A}^{\mathrm{H}} = (\boldsymbol{B} + \mathrm{i}\boldsymbol{C})(\boldsymbol{B}^{\mathrm{T}} - \mathrm{i}\boldsymbol{C}^{\mathrm{T}}) = \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{C}\boldsymbol{C}^{\mathrm{T}} + \mathrm{i}(\boldsymbol{C}\boldsymbol{B}^{\mathrm{T}} - \boldsymbol{B}\boldsymbol{C}^{\mathrm{T}})$$

于是有

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} + \mathrm{i}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{C} - \boldsymbol{C}^{\mathrm{T}}\boldsymbol{B}) = \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{C}\boldsymbol{C}^{\mathrm{T}} + \mathrm{i}(\boldsymbol{C}\boldsymbol{B}^{\mathrm{T}} - \boldsymbol{B}\boldsymbol{C}^{\mathrm{T}})$$

即

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} - \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} - \boldsymbol{C}\boldsymbol{C}^{\mathrm{T}} = \mathrm{i}(\boldsymbol{C}\boldsymbol{B}^{\mathrm{T}} - \boldsymbol{B}\boldsymbol{C}^{\mathrm{T}} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}}\boldsymbol{C})$$

注意到 $\boldsymbol{B}$, $\boldsymbol{C}$ 是实矩阵，故由上式得

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} - \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} - \boldsymbol{C}\boldsymbol{C}^{\mathrm{T}} = \boldsymbol{0}, \quad \boldsymbol{C}\boldsymbol{B}^{\mathrm{T}} - \boldsymbol{B}\boldsymbol{C}^{\mathrm{T}} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}}\boldsymbol{C} = \boldsymbol{0}$$

所以

$$\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} = \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{C}\boldsymbol{C}^{\mathrm{T}}, \quad \boldsymbol{C}^{\mathrm{T}}\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}}\boldsymbol{C} = \boldsymbol{B}\boldsymbol{C}^{\mathrm{T}} - \boldsymbol{C}\boldsymbol{B}^{\mathrm{T}} \qquad ①$$

再计算

$$\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} = \begin{pmatrix} \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} & \boldsymbol{C}^{\mathrm{T}}\boldsymbol{B} - \boldsymbol{B}^{\mathrm{T}}\boldsymbol{C} \\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{C} - \boldsymbol{C}^{\mathrm{T}}\boldsymbol{B} & \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} + \boldsymbol{C}^{\mathrm{T}}\boldsymbol{C} \end{pmatrix}$$

$$\boldsymbol{R}\boldsymbol{R}^{\mathrm{T}} = \begin{pmatrix} \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{C}\boldsymbol{C}^{\mathrm{T}} & \boldsymbol{B}\boldsymbol{C}^{\mathrm{T}} - \boldsymbol{C}\boldsymbol{B}^{\mathrm{T}} \\ \boldsymbol{C}\boldsymbol{B}^{\mathrm{T}} - \boldsymbol{B}\boldsymbol{C}^{\mathrm{T}} & \boldsymbol{B}\boldsymbol{B}^{\mathrm{T}} + \boldsymbol{C}\boldsymbol{C}^{\mathrm{T}} \end{pmatrix}$$

根据式①立即可知

$$\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R} = \boldsymbol{R}\boldsymbol{R}^{\mathrm{T}}$$

(2) 当 $\boldsymbol{A}$ 是 Hermite 时，由 $\boldsymbol{A}^{\mathrm{H}} = \boldsymbol{A}$ 可得

$$\boldsymbol{B}^{\mathrm{T}} = \boldsymbol{B}, \qquad \boldsymbol{C}^{\mathrm{T}} = -\boldsymbol{C}$$

所以

$$\boldsymbol{R}^{\mathrm{T}}=\begin{bmatrix}\boldsymbol{B}^{\mathrm{T}} & \boldsymbol{C}^{\mathrm{T}}\\ -\boldsymbol{C}^{\mathrm{T}} & \boldsymbol{B}^{\mathrm{T}}\end{bmatrix}=\begin{bmatrix}\boldsymbol{B} & -\boldsymbol{C}\\ \boldsymbol{C} & \boldsymbol{B}\end{bmatrix}=\boldsymbol{R}$$

(3) 当 $\boldsymbol{A}$ 是 Hermite 正定矩阵时,对任意 $\boldsymbol{0}\neq\boldsymbol{x}\in\mathbf{C}^n$,有

$$\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{x}^{\mathrm{H}}(\boldsymbol{B}+\mathrm{i}\boldsymbol{C})\boldsymbol{x}>0$$

特别地,当 $\boldsymbol{x}$ 是实向量时,由 $\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}+\mathrm{i}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{x}>0$ 可知

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}>0,\qquad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{C}\boldsymbol{x}=0$$

从而 $\boldsymbol{B}$ 是实正定矩阵. 任取 $\boldsymbol{y}\in\mathbf{R}^{2n}$, $\boldsymbol{y}=\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix}\neq\boldsymbol{0}$, $\boldsymbol{y}_1$, $\boldsymbol{y}_2\in\mathbf{R}^n$ 至少有一个是非零向量. 计算

$$\begin{aligned}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y}&=(\boldsymbol{y}_1^{\mathrm{T}},\ \boldsymbol{y}_2^{\mathrm{T}})\begin{bmatrix}\boldsymbol{B} & \boldsymbol{C}\\ \boldsymbol{C} & \boldsymbol{B}\end{bmatrix}\begin{bmatrix}\boldsymbol{y}_1\\ \boldsymbol{y}_2\end{bmatrix}\\ &=\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{B}\boldsymbol{y}_1+\boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{B}\boldsymbol{y}_2+\boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y}_1+\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y}_2\end{aligned}\tag{②}$$

由 $\boldsymbol{B}$ 的正定性可知

$$\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{B}\boldsymbol{y}_1+\boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{B}\boldsymbol{y}_2>0$$

又因

$$\begin{aligned}\boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y}_1+\boldsymbol{y}_1^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y}_2&=\boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y}_1-(\boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y}_1)^{\mathrm{T}}\\ &=\boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y}_1-\boldsymbol{y}_2^{\mathrm{T}}\boldsymbol{C}\boldsymbol{y}_1=0\end{aligned}$$

所以,根据式②得

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{y}>0$$

即 $\boldsymbol{W}$ 是实正定矩阵.

(4) 由于

$$\begin{aligned}\boldsymbol{E}=\boldsymbol{A}^{\mathrm{H}}\boldsymbol{A}&=(\boldsymbol{B}^{\mathrm{T}}-\mathrm{i}\boldsymbol{C}^{\mathrm{T}})(\boldsymbol{B}+\mathrm{i}\boldsymbol{C})\\ &=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}+\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}+\mathrm{i}(\boldsymbol{B}^{\mathrm{T}}\boldsymbol{C}-\boldsymbol{C}^{\mathrm{T}}\boldsymbol{B})\end{aligned}$$

故

$$\begin{cases}\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}+\boldsymbol{C}^{\mathrm{T}}\boldsymbol{C}=\boldsymbol{E}\\ \boldsymbol{B}^{\mathrm{T}}\boldsymbol{C}-\boldsymbol{C}^{\mathrm{T}}\boldsymbol{B}=\boldsymbol{0}\end{cases}$$

由此可得

$$R^{\mathrm{T}}R=\begin{pmatrix}B^{\mathrm{T}}B+C^{\mathrm{T}}C & C^{\mathrm{T}}B-B^{\mathrm{T}}C\\ B^{\mathrm{T}}C-C^{\mathrm{T}}B & B^{\mathrm{T}}B+C^{\mathrm{T}}C\end{pmatrix}=\begin{pmatrix}E & \\ & E\end{pmatrix}$$

即知 $\boldsymbol{R}$ 是正交矩阵.

5. $A=U\Lambda U^{\mathrm{H}}$, $U$ 为酉矩阵, $\Lambda=\mathrm{diag}\{\lambda_1,\ \cdots,\ \lambda_n\}$, $\lambda_i$ 为 $A$ 的特征值. 于是由 $AA^{\mathrm{H}}=U\Lambda^2U^{\mathrm{H}}$ 知 $A$ 的奇异值为

$$S_i=|\lambda_i|,\quad i=1,\ \cdots,\ n$$

6. 由 5 题结论知.

7. $$A=\begin{pmatrix}\frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & 0 & -\frac{2}{\sqrt{6}}\\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}}\end{pmatrix}\begin{pmatrix}\sqrt{15} & 0\\ 0 & 0\\ 0 & 0\end{pmatrix}\begin{pmatrix}\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}}\\ \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}}\end{pmatrix}.$$

8. $A=\begin{pmatrix}1\\0\end{pmatrix}(1,\ 0,\ 0)$ 是 $A$ 的秩分解,由式(4.5.5)易知

$$A^{+}=B$$

又根据定理 4.5.2(i)得

$$B^{+}=(A^{+})^{+}=A$$

当然,如果已证 $A^{+}=B$, 则 $B^{+}=A$ 就不必再证明了,因为由 Moore-Penrose 逆的定义可知,加号逆是对称的.

9. 由于 $A$, $B$ 分别是列满秩和行满秩矩阵,根据推论 4.5.1 可知

$$A^{+}=(A^{\mathrm{H}}A)^{-1}A^{\mathrm{H}},\qquad B^{+}=B^{\mathrm{H}}(BB^{\mathrm{H}})^{-1}$$

注意到 $AB$ 自身构成秩分解,故

$$(AB)^{+}=B^{\mathrm{H}}(BB^{\mathrm{H}})^{-1}(A^{\mathrm{H}}A)^{-1}A^{\mathrm{H}}=B^{+}A^{+}$$

10. $A^{+}=\dfrac{1}{14}\begin{pmatrix}3 & 5 & -1\\ 2 & 1 & 4\end{pmatrix}$.

## 习　题　5

1. $\|\alpha\|_1=7+\sqrt{2}$, $\|\alpha\|_2=\sqrt{23}$, $\|\alpha\|_\infty=4$.

2. (1) 因为

$$\|\boldsymbol{\alpha}\|_1^2=(|x_1|+\cdots+|x_n|)^2$$
$$\geqslant|x_1|^2+|x_2|^2+\cdots+|x_n|^2=\|\alpha\|_2^2$$

所以

$$\|\boldsymbol{\alpha}\|_2\leqslant\|\boldsymbol{\alpha}\|_1$$

又因

$$n\|\boldsymbol{\alpha}\|_2^2-\|\boldsymbol{\alpha}\|_1^2=n(|x_1|^2+\cdots+|x_n|^2)-(|x_1|+\cdots+|x_n|)^2$$
$$=\sum_{1\leqslant i<j\leqslant n}(|x_i|-|x_j|)^2\geqslant 0$$

因此得

$$\|\boldsymbol{\alpha}\|_1\leqslant\sqrt{n}\|\boldsymbol{\alpha}\|_2$$

(2) 设 $\max\limits_{1\leqslant i\leqslant n}|x_i|=|x_k|$，则

$$\|\boldsymbol{\alpha}\|_\infty=|x_k|$$

另一方面又有

$$\|\boldsymbol{\alpha}\|_1=|x_1|+\cdots+|x_n|\geqslant|x_k|=\|\boldsymbol{\alpha}\|_\infty$$
$$\|\boldsymbol{\alpha}\|_1=|x_1|+\cdots+|x_n|\leqslant n|x_k|=n\|\boldsymbol{\alpha}\|_\infty$$

(3) 由下列不等式便得证.

由(2)知

$$\max_{1\leqslant i\leqslant n}|x_i|=|x_k$$
$$\|\boldsymbol{\alpha}\|_\infty^2=|x_k|^2\leqslant\sum|x_i|^2=\|\alpha\|_2^2$$
$$=\sum|x_i|^2\leqslant n|x_k|^2=n\|\boldsymbol{\alpha}\|_\infty^2$$

3. (1) 当 $\boldsymbol{\alpha}\neq\mathbf{0}$ 时，由 $\|\boldsymbol{\alpha}\|_a>0$，$\|\boldsymbol{\alpha}\|_b>0$ 知

$$\max\{\|\boldsymbol{\alpha}\|_a,\|\boldsymbol{\alpha}\|_b\}>0$$

即正定性成立.

对任意数 $k$，有

$$\|k\boldsymbol{\alpha}\|=\max\{\|k\boldsymbol{\alpha}\|_a,\|k\boldsymbol{\alpha}\|_b\}$$
$$=\max\{|k|\|\boldsymbol{\alpha}\|_a,|k|\|\boldsymbol{\alpha}\|_b\}$$
$$=|k|\max\{\|\boldsymbol{\alpha}\|_a,\|\boldsymbol{\alpha}\|_b\}$$

齐次性成立.

对任意 $\boldsymbol{\beta} \in \mathbf{C}^n$，有

$$\begin{aligned}\|\boldsymbol{\alpha}+\boldsymbol{\beta}\| &= \max\{\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_a, \|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_b\} \\ &\leqslant \max\{\|\boldsymbol{\alpha}\|_a+\|\boldsymbol{\beta}\|_a, \|\boldsymbol{\alpha}\|_b+\|\boldsymbol{\beta}\|_b\} \\ &\leqslant \max\{\|\boldsymbol{\alpha}\|_a, \|\boldsymbol{\alpha}\|_b\}+\max\{\|\boldsymbol{\beta}\|_a, \|\boldsymbol{\beta}\|_b\} \\ &= \|\boldsymbol{\alpha}\|+\|\boldsymbol{\beta}\|\end{aligned}$$

即满足三角不等式. 故是 $\mathbf{C}^n$ 上的向量范数.

(2) 当 $\boldsymbol{\alpha}=\mathbf{0}$ 时

$$\|\boldsymbol{\alpha}\| = k_1 0 + k_2 0 = 0$$

当 $\|\boldsymbol{\alpha}\| \neq 0$ 时，由 $\|\boldsymbol{\alpha}\|_a > 0$，$\|\boldsymbol{\alpha}\|_b > 0$ 知

$$\|\boldsymbol{\alpha}\| = k_1\|\boldsymbol{\alpha}\|_a + k_2\|\boldsymbol{\alpha}\|_b > 0$$

故正定性成立.

对任意数 $k$ 有

$$\begin{aligned}\|k\boldsymbol{\alpha}\| &= k_1\|k\boldsymbol{\alpha}\|_a + k_2\|k\boldsymbol{\alpha}\|_b \\ &= k_1|k|\,\|\boldsymbol{\alpha}\|_a + k_2|k|\,\|\boldsymbol{\alpha}\|_b \\ &= |k|(k_1\|\boldsymbol{\alpha}\|_a + k_2\|\boldsymbol{\alpha}\|_b) \\ &= |k|\,\|\boldsymbol{\alpha}\|\end{aligned}$$

齐次性成立. 又对任意 $\boldsymbol{\beta} \in \mathbf{C}^n$，有

$$\begin{aligned}\|\boldsymbol{\alpha}+\boldsymbol{\beta}\| &= k_1\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_a + k_2\|\boldsymbol{\alpha}+\boldsymbol{\beta}\|_b \\ &\leqslant k_1(\|\boldsymbol{\alpha}\|_a+\|\boldsymbol{\beta}\|_a)+k_2(\|\boldsymbol{\alpha}\|_b+\|\boldsymbol{\beta}\|_b) \\ &= (k_1\|\boldsymbol{\alpha}\|_a+k_2\|\boldsymbol{\alpha}\|_b)+(k_1\|\boldsymbol{\beta}\|_a+k_2\|\boldsymbol{\beta}\|_b) \\ &= \|\boldsymbol{\alpha}\|+\|\boldsymbol{\beta}\|\end{aligned}$$

三角不等式成立，从而是 $\mathbf{C}^n$ 上的向量范数.

4. (1) 当 $\mathbf{A}=\mathbf{0}$ 时

$$\|\mathbf{A}\| = 0$$

当 $\mathbf{A} \neq \mathbf{0}$ 时

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} \neq \mathbf{0}$$

故

$$\|A\| = \|P^{-1}AP\|_m > 0$$

(2) 对任意 $k \in \mathbf{C}$，有

$$\begin{aligned}\|kA\| &= \|P^{-1}(kA)P\|_m = |k|\ \|P^{-1}AP\|_m \\ &= |k|\ \|A\|\end{aligned}$$

(3)

$$\begin{aligned}\|A+B\| &= \|P^{-1}(A+B)P\|_m \\ &= \|P^{-1}AP + P^{-1}BP\|_m \\ &\leqslant \|P^{-1}AP\|_m + \|P^{-1}BP\|_m \\ &= \|A\| + \|B\|\end{aligned}$$

(4)

$$\begin{aligned}\|AB\| &= \|P^{-1}(AB)P\|_m \\ &= \|(P^{-1}AP)(P^{-1}BP)\|_m \\ &\leqslant \|P^{-1}AP\|_m \|P^{-1}BP\|_m \\ &= \|A\|\ \|B\|\end{aligned}$$

从而 $\|A\|$ 是 $\mathbf{C}^{n\times n}$ 上的范数.

5. $\|A\|_1 = 12$，$\|A\|_\infty = 14$，$\|Ax\|_1 = 9+\sqrt{10}+\sqrt{53}$，$\|Ax\|_\infty = \sqrt{53}$.

6. $\|A\|_{m_1} = 18+\sqrt{2}$，$\|A\|_F = \sqrt{66}$，$\|A\|_{m_\infty} = 15$，$\|A\|_1 = 7+\sqrt{2}$，$\|A\|_\infty = 9$.

7. 用定义证明.

8. 设 $A = (a_{ij})_{m\times n}$，$\alpha = (x_1, x_2, \cdots, x_n)^T$，则

$$A = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}E_{ij}, \qquad E_{ij}\alpha = (0, \cdots, 0, x_j, 0, \cdots, 0)^T$$

$$\|E_{ij}\alpha\|_p \leqslant \|\alpha\|_p$$

故

$$\begin{aligned}\|A\alpha\|_p &= \left\|\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}E_{ij}\alpha\right\|_p \\ &\leqslant \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|\ \|E_{ij}\alpha\|_p \\ &\leqslant \sum_{i=1}^{m}\sum_{j=1}^{n} |a_{ij}|\ \|\alpha\|_p \\ &= \|A\|_{m_1}\ \|\alpha\|_p\end{aligned}$$

即矩阵范数 $\|A\|_{m_1}$ 与向量范数 $p$-范数相容.

9. 事实上

$$\|UA\|_F^2 = \mathrm{tr}[(UA)^H(UA)] = \mathrm{tr}[A^H(U^HU)A]$$
$$= \mathrm{tr}(A^HA) = \|A\|_F^2$$

即

$$\|UA\|_F = \|A\|_F$$

又因为

$$\|A\|_F = \|A^H\|_F$$

故有

$$\|AU\|_F = \|U^HA^H\|_F = \|A^H\|_F = \|A\|_F$$

知结论成立.

10. $\mathrm{Cond}(A_1) = \dfrac{259}{2}$, $\mathrm{Cond}(A)_\infty = \dfrac{611}{4}$.

11. $\hat{x} = A^+ b = \dfrac{1}{10}\begin{bmatrix}1\\2\\3\end{bmatrix}$.

## 习 题 6

1. (1) 可以求得 $A$ 的特征值为

$$\lambda_1 = \frac{5}{6}, \qquad \lambda_2 = -\frac{1}{2}$$

于是

$$\rho(A) = \frac{5}{6} < 1$$

故 $A$ 是收敛矩阵.

(2) 因为 $\|A\|_1 = 0.8 < 1$, 所以 $A$ 是收敛矩阵.

2. (1) 矩阵级数 $\sum\limits_{k=0}^{+\infty} A^{(k)}$ 收敛, 其和

$$S = \lim_{N\to+\infty} S^{(N)} = \begin{bmatrix}2 & \frac{3}{2}\\ 1 & 0\end{bmatrix}$$

(2) $\sum\limits_{k=0}^{+\infty}\begin{pmatrix}0.1 & 0.7\\ 0.3 & 0.6\end{pmatrix}^k = (E-A)^{-1} = \dfrac{1}{0.15}\begin{pmatrix}0.4 & 0.7\\ 0.3 & 0.9\end{pmatrix}$.

3. $e^{A}=\begin{bmatrix} e^2 & 0 & 0 \\ e^2 & 0 & e^2 \\ e^2 & -e^2 & 2e^2 \end{bmatrix}$,

$$e^{At}=\begin{bmatrix} e^{2t} & 0 & 0 \\ te^{2t} & (1-t)e^{2t} & te^{2t} \\ te^{2t} & -te^{2t} & (1+t)e^{2t} \end{bmatrix},$$

$$\sin A=\begin{bmatrix} \sin 2 & 0 & 0 \\ \cos 2 & \sin 2-\cos 2 & \cos 2 \\ \cos 2 & -\cos 2 & \sin 2+\cos 2 \end{bmatrix}.$$

4. $e^{At}=\begin{bmatrix} e^{2t} & 12e^{t}-12e^{2t}+13te^{2t} & -4e^{t}+4e^{2t} \\ 0 & e^{2t} & 0 \\ 0 & -3e^{t}+3e^{2t} & e^{t} \end{bmatrix}$,

$$\sin At=\begin{bmatrix} \sin 2t & 12\sin t-12\sin 2t+13t\cos 2t & -4\sin t+4\sin 2t \\ 0 & \sin 2t & 0 \\ 0 & -3\sin t+3\sin 2t & \sin t \end{bmatrix}.$$

5. $\ln A=\begin{bmatrix} \ln 1 & 0 & 0 & 0 \\ \frac{1}{1} & \ln 1 & 0 & 0 \\ \frac{1}{2!}\left(-\frac{1}{1^2}\right) & \frac{1}{1} & \ln 1 & 0 \\ \frac{1}{3!}\left(2\times\frac{1}{1^3}\right) & \frac{1}{2!}\left(-\frac{1}{1^2}\right) & \frac{1}{1} & \ln 1 \end{bmatrix}$

$$=\begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 & 0 \\ \frac{1}{3} & -\frac{1}{2} & 1 & 0 \end{bmatrix}.$$

6. (1) $\frac{d}{dt}A(t)=\begin{bmatrix} 2e^{2t} & (1+t)e^{t} & 0 \\ -e^{-t} & 4e^{2t} & 0 \\ 3 & 0 & 0 \end{bmatrix}$,

$$\frac{d}{dt}\,|\,A(t)\,|=-6(1+2t)e^{2t};$$

(2) $\int A(t)\mathrm{d}t=\begin{bmatrix}\frac{1}{2}\mathrm{e}^{2t}+C_{11} & (t-1)\mathrm{e}^{t}+C_{12} & t+C_{13}\\ -\mathrm{e}^{-t}+C_{21} & \mathrm{e}^{2t}+C_{22} & C_{23}\\ \frac{3}{2}t^2+C_{31} & C_{32} & C_{33}\end{bmatrix}$,

$$\int_0^1 A(t)\mathrm{d}t=\begin{bmatrix}\frac{1}{2}(\mathrm{e}^2-1) & 1 & 1\\ 1-\mathrm{e}^{-1} & \mathrm{e}^2-1 & 0\\ \frac{3}{2} & 0 & 0\end{bmatrix}.$$

7. 首先易证矩阵序列与级数下列性质成立:

(1) 若矩阵序列 $A^{(k)}\to A$,则$(A^{\mathrm{T}})^{(k)}\to A^{\mathrm{T}}$, $\overline{A}^{(k)}\to\overline{A}$;

(2) 若矩阵级数 $\sum_{k=0}^{+\infty}C_kA^k$ 收敛,则

$$\left(\sum_{k=0}^{+\infty}C_kA^k\right)^{\mathrm{T}}=\sum_{k=0}^{+\infty}C_k(A^{\mathrm{T}})^k$$

由此得

$$(\mathrm{e}^{A})^{\mathrm{T}}=\left(\sum_{k=0}^{+\infty}\frac{A^{(k)}}{k!}\right)^{\mathrm{T}}=\sum_{k=0}^{+\infty}\frac{(A^{\mathrm{T}})^{(k)}}{k!}=\mathrm{e}^{A^{\mathrm{T}}}$$

所以

①当 $A$ 是实反对称阵时,有

$$\mathrm{e}^{A}\cdot(\mathrm{e}^{A})^{\mathrm{T}}=\mathrm{e}^{A}\cdot\mathrm{e}^{A^{\mathrm{T}}}=\mathrm{e}^{A}\cdot\mathrm{e}^{-A}=\mathrm{e}^{0}=E$$

②当 $A$ 为 Hermite 矩阵时,即 $A^{\mathrm{H}}=A$,有

$$(\mathrm{e}^{\mathrm{i}A})^{\mathrm{H}}\cdot\mathrm{e}^{\mathrm{i}A}=\mathrm{e}^{(\mathrm{i}A)^{\mathrm{H}}}\cdot\mathrm{e}^{\mathrm{i}A}=\mathrm{e}^{-\mathrm{i}A^{\mathrm{H}}}\cdot\mathrm{e}^{\mathrm{i}A}$$
$$=\mathrm{e}^{-\mathrm{i}A+\mathrm{i}A}=\mathrm{e}^{0}=E$$

故 $\mathrm{e}^{\mathrm{i}A}$为酉矩阵.

8. $\det A=\sum_{j=1}^{n}x_{ij}A_{ij}$,其中 $A_{ij}$是 $x_{ij}$ 的代数余子式,$A_{ij}$不含 $x_{ij}$,故

$$\frac{\partial}{\partial x_{ij}}\det A=A_{ij}$$

所以

$$\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{A}}\det\boldsymbol{A}=(A_{ij})_{n\times n}=(\boldsymbol{A}^{*})^{\mathrm{T}}$$

$$=[(\det\boldsymbol{A})\boldsymbol{A}^{-1}]^{\mathrm{T}}=\det\boldsymbol{A}(\boldsymbol{A}^{-1})^{\mathrm{T}}$$

9. $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}-b.$

10. $\dfrac{\mathrm{d}f}{\mathrm{d}\boldsymbol{A}}=2\boldsymbol{A}.$

11. $\dfrac{\mathrm{d}F(\boldsymbol{\alpha})}{\mathrm{d}\boldsymbol{\alpha}}=(a_{11},\cdots,a_{m1},\cdots,a_{1n},\cdots,a_{mn})^{\mathrm{T}}.$

12. $\dfrac{\mathrm{d}F(\boldsymbol{\alpha})}{\mathrm{d}\boldsymbol{\alpha}}=\boldsymbol{A}.$

13. $\boldsymbol{x}(t)=\begin{pmatrix}\frac{1}{2}\mathrm{e}^{3t}+\frac{1}{2}\mathrm{e}^{-t}\\ \frac{1}{4}\mathrm{e}^{3t}-\frac{1}{4}\mathrm{e}^{-t}\\ \frac{1}{2}\mathrm{e}^{3t}-\frac{1}{2}\mathrm{e}^{-t}\end{pmatrix}.$

14. $\boldsymbol{x}(t)=\begin{pmatrix}(2-t)\mathrm{e}^{t}-1\\ (t-1)\mathrm{e}^{t}+1\\ (3-2t)\mathrm{e}^{t}-2\end{pmatrix}.$

15. 由

$$\boldsymbol{A}=\begin{pmatrix}1&1&0\\0&1&0\\0&1&1\end{pmatrix},\quad \boldsymbol{B}=\begin{pmatrix}0&1\\1&0\\0&1\end{pmatrix}$$

得

$$\boldsymbol{W}_C=(\boldsymbol{B},\boldsymbol{A}\boldsymbol{B},\boldsymbol{A}^2\boldsymbol{B})=\begin{pmatrix}0&1&1&1&2&1\\1&0&1&0&1&0\\0&1&1&1&2&1\end{pmatrix}$$

$$|\boldsymbol{W}_C|=2<3=n$$

故系统是不能控的.

16. 由

$$\boldsymbol{A}=\begin{pmatrix}1&0\\0&1\end{pmatrix},\quad \boldsymbol{C}=\begin{pmatrix}1&1\\2&2\end{pmatrix}$$

得

$$W_0=\begin{pmatrix}C\\CA\end{pmatrix}=\begin{pmatrix}0&0\\2&0\\0&0\\2&0\end{pmatrix}$$

$$|W_0|=1<2=n$$

故系统是不能观测的.

## 习 题 7

1. 取 $A=\begin{pmatrix}1&1\\1&-1\end{pmatrix}$, $B=\begin{pmatrix}0&-0.4\\0.9&1\end{pmatrix}$.

2. $A$ 的三个盖尔圆是

$$|z|\leqslant 1,\quad |z-5|\leqslant 2,\quad |z-9|\leqslant 3$$

令 $D=\mathrm{diag}(2,1,1)$,则 $B=DAD^{-1}$ 的三个盖尔圆是

$$|z|\leqslant 2,\quad |z-5|\leqslant \frac{3}{2},\quad |z-9|\leqslant 2$$

因为 $B$ 的特征值就是 $A$ 的特征值,而 $B$ 的盖尔圆是彼此隔离的,所以在以上每个闭圆盘中恰有一个特征值. 又因圆心都在实轴上,若有复特征值,则必成对共轭出现,也就是在某一圆中有两个特征值,这与圆盘定理的结论矛盾. 即 $A$ 的特征值全为实数且满足条件

$$-2\leqslant\lambda_1\leqslant 2,\quad 3.5\leqslant\lambda_2\leqslant 6.5,\quad 7\leqslant\lambda_3\leqslant 11$$

3. $A^HA$ 是半正定矩阵,其特征值是非负实数. 因而可以设

$$\mu_1\geqslant\mu_2\geqslant\cdots\geqslant\mu_r\geqslant 0=\mu_{r+1}=\cdots=\mu_n$$

于是 $A$ 的奇异值分解为

$$A=U\left(\begin{array}{ccc|c}\sqrt{\mu_1}&&&\\&\ddots&&\mathbf{0}\\&&\sqrt{\mu_r}&\\\hline&\mathbf{0}&&\mathbf{0}\end{array}\right)V^H$$

根据定理 7.2.1 立即得到

$$\sum_{i=1}^{n}\mu_i=\sum_{i=1}^{r}\mu_i=\|\boldsymbol{A}\|_F^2\geqslant\sum_{i=1}^{n}|\lambda_i|^2$$

4. 作 $n$ 阶矩阵 $\boldsymbol{A}$，

$$\boldsymbol{A}=\begin{pmatrix}0 & 0 & \cdots & 0 & -\dfrac{a_n}{a_0}\\ 1 & 0 & \cdots & 0 & -\dfrac{a_{n-1}}{a_0}\\ 0 & 1 & \cdots & 0 & -\dfrac{a_{n-2}}{a_0}\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & 0 & -\dfrac{a_2}{a_0}\\ 0 & 0 & \cdots & 1 & -\dfrac{a_1}{a_0}\end{pmatrix}$$

求出其 $\boldsymbol{A}$ 的特征多项式

$$|\lambda\boldsymbol{I}-\boldsymbol{A}|=\begin{vmatrix}\lambda & 0 & \cdots & 0 & \dfrac{a_n}{a_0}\\ -1 & \lambda & \cdots & 0 & \dfrac{a_{n-1}}{a_0}\\ 0 & -1 & \cdots & 0 & \dfrac{a_{n-2}}{a_0}\\ \vdots & \vdots & & \vdots & \vdots\\ 0 & 0 & \cdots & \lambda & \dfrac{a_2}{a_0}\\ 0 & 0 & \cdots & -1 & \lambda+\dfrac{a_1}{a_0}\end{vmatrix}$$

$$=\lambda^n+\frac{a_1}{a_0}\lambda^{n-1}+\frac{a_2}{a_0}\lambda^{n-2}+\cdots+\frac{a_n}{a_0}$$

$$=\frac{1}{a_0}f(\lambda)$$

所以，$f(z)$的零点是矩阵 $\boldsymbol{A}$ 的特征值. 于是根据 Geršgorin 圆盘定理，$f(z)$的

零点必在下列 $n$ 个闭圆盘中

$$|z| \leqslant \left|\frac{a_n}{a_0}\right|, \quad |z| \leqslant 1+\left|\frac{a_k}{a_0}\right|, \quad \left|z+\frac{a_1}{a_0}\right| \leqslant 1$$

其中 $k=2, \cdots, n-1$，所以

$$|z_0| \leqslant 1+\max_{1\leqslant k\leqslant n}\left|\frac{a_k}{a_0}\right|$$

5. (1) 设 $R(\boldsymbol{x})=(\boldsymbol{H}\boldsymbol{x}, \boldsymbol{x})$，$\|\boldsymbol{x}\|_2=1$，则由定理 7.3.1 知，$\lambda_1 \leqslant R(\boldsymbol{x}) \leqslant \lambda_n$，其中 $\|\boldsymbol{x}\|_2=1$. 取 $\boldsymbol{x}=\boldsymbol{e}_j=(0, \cdots, 0, \overset{(j)}{1}, 0, \cdots, 0)^{\mathrm{T}}$，则 $\|\boldsymbol{x}\|_2=1$，且

$$\lambda_1 \leqslant R(\boldsymbol{e}_j)=h_{jj} \leqslant \lambda_n$$

(2) 利用(1)的结果，取 $\boldsymbol{x}=n^{-\frac{1}{2}}(1, 1, \cdots, 1)^{\mathrm{T}}$.

6. 显然，$V_{i+1}^{\perp}=\mathrm{Span}\{\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_i\}$. 任取 $\boldsymbol{x} \in \boldsymbol{V}_{i+1}^{\perp}$，$\boldsymbol{x} \neq \boldsymbol{0}$，则存在常数 $k_1$，$k_2$，$\cdots$，$k_i$ 不全为 0，使得 $\boldsymbol{x}=\sum_{j=1}^{i} k_j \boldsymbol{x}_j$. 令 $\boldsymbol{y}=(k_1, k_2, \cdots, k_i, 0, \cdots, 0)^{\mathrm{T}} \in \mathbf{C}^n$，则有 $\boldsymbol{x}=\boldsymbol{U}\boldsymbol{y}$，并且

$$(\boldsymbol{H}\boldsymbol{x}, \boldsymbol{x})=(\boldsymbol{H}\boldsymbol{U}\boldsymbol{y}, \boldsymbol{U}\boldsymbol{y})=\boldsymbol{y}^{\mathrm{H}}\boldsymbol{\Lambda}\boldsymbol{y}=\sum_{j=1}^{i}\lambda_j |k_j|^2, \quad (\boldsymbol{x}, \boldsymbol{x})=\sum_{j=1}^{i}|k_j|^2$$

于是有

$$R(\boldsymbol{x})=\frac{(\boldsymbol{H}\boldsymbol{x}, \boldsymbol{x})}{(\boldsymbol{x}, \boldsymbol{x})}=\frac{\lambda_1 |k_1|^2+\lambda_2 |k_2|^2+\cdots+\lambda_i |k_i|^2}{|k_1|^2+|k_2|^2+\cdots+|k_i|^2} \leqslant \lambda_i$$

因为 $R(\boldsymbol{x}_i)=\lambda_i$，所以

$$\lambda_i=\max_{0\neq \boldsymbol{x}\in V_{i+1}^{\perp}} R(\boldsymbol{x})$$

7. 将引理 7.3.1 证明中的子空间 $\hat{W}_j$ 换成 $V_{n-j+1}$，然后作类似的证明.

8. $\boldsymbol{A}_n=\begin{pmatrix}\boldsymbol{A}_{n-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}$，对任意的 $\boldsymbol{x}=(x_1, x_2, \cdots, x_{n-1})^{\mathrm{T}} \in \mathbf{C}^{n-1}$，$(\boldsymbol{x}, \boldsymbol{x})=1$，由定理 7.3.1 知 $\mu=\max R(\boldsymbol{x})$. 令 $\hat{\boldsymbol{x}}=(x_1, x_2, \cdots, x_{n-1}, 0)^{\mathrm{T}}$，则

$$\hat{\boldsymbol{x}}^{\mathrm{H}}\boldsymbol{A}_n\hat{\boldsymbol{x}}=(\boldsymbol{x}^{\mathrm{H}}, 0)\begin{pmatrix}\boldsymbol{A}_{n-1} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{x} \\ \boldsymbol{0}\end{pmatrix}=(\boldsymbol{A}_{n-1}\boldsymbol{x}, \boldsymbol{x})=\boldsymbol{x}^{\mathrm{H}}\boldsymbol{A}_{n-1}\boldsymbol{x}=R(\boldsymbol{x})$$

上式表明 $\underset{0\neq x\in \mathbf{C}^{n-1}}{R(x)} \geqslant \min\limits_{0\neq \boldsymbol{x}\in \mathbf{C}^n}(\boldsymbol{A}_n\boldsymbol{x}, \boldsymbol{x})=\lambda_1$，故

$$\mu=\max_{0\neq \boldsymbol{x}\in \mathbf{C}^{n-1}} R(\boldsymbol{x}) \geqslant \min_{0\neq \boldsymbol{x}\in \mathbf{C}^n}(\boldsymbol{A}_n\boldsymbol{x}, \boldsymbol{x})=\lambda_1$$

另一方面，对任意 $n-1$-维子空间 $W_{n-1}$ 及如上所设的 $\boldsymbol{x}$ 与 $\hat{\boldsymbol{x}}$，有

$$\mu = \max_{0\neq x\in \mathbf{C}^{n-1}} R(\boldsymbol{x}) = \max_{0\neq \hat{x}\in \mathbf{C}^{n}} (\boldsymbol{A}_n \hat{\boldsymbol{x}},\ \hat{\boldsymbol{x}})$$

于是由定理 7.3.3 得

$$\mu \leqslant \min_{W_{n-1}} \max_{0\neq x\in W_{n-1}} (\boldsymbol{A}_n \boldsymbol{x},\ \boldsymbol{x}) = \lambda_2$$

9. 因 $\boldsymbol{A}^k > 0$，所以 $\rho(\boldsymbol{A}^k) > 0$. 根据 $\rho(\boldsymbol{A}^k) = \rho(\boldsymbol{A})^k$ 知 $\rho(\boldsymbol{A}) > 0$.
10. 假设 $\rho(\boldsymbol{A}) = 0$. 由非负矩阵的 Perron 定理知，$\boldsymbol{A}$ 的特征值都为 0. 根据条件，存在正的特征向量 $\boldsymbol{y}$，于是有 $\boldsymbol{A}\boldsymbol{y} = 0$. 但这与 $\boldsymbol{A}\boldsymbol{y} > 0$ 矛盾.
11. 首先对于矩阵的任意范数 $\|\cdot\|$，有 $\rho(\boldsymbol{A}) \leqslant \|\boldsymbol{A}\|$. 另一方面，如果 $\boldsymbol{A}$ 的各个行和是常数，则向量 $\boldsymbol{x} = (1,\ 1,\ \cdots,\ 1)^{\mathrm{T}}$ 为 $\boldsymbol{A}$ 的对应于特征值 $\|\boldsymbol{A}\|_\infty$ 的特征向量，所以有 $\rho(\boldsymbol{A}) = \|\boldsymbol{A}\|_\infty$. 对于矩阵 $\boldsymbol{A}^{\mathrm{T}}$ 应用同样的讨论可以得到 $\rho(\boldsymbol{A}) = \|\boldsymbol{A}\|_1$.
12. $\boldsymbol{y}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x} - \rho(\boldsymbol{A})\boldsymbol{x}) = (\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y})^{\mathrm{T}}\boldsymbol{x} - \rho(\boldsymbol{A})\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x} = 0$，又 $\boldsymbol{y} > 0$ 且 $\boldsymbol{A}\boldsymbol{x} \geqslant \rho(\boldsymbol{A})\boldsymbol{x}$，所以 $\boldsymbol{A}\boldsymbol{x} = \rho(\boldsymbol{A})\boldsymbol{x}$.

# 参考文献


Horn R A，等. 2005. 矩阵分析. 杨奇，译. 北京：机械工业出版社

北京大学数学系. 2003. 高等代数. 第三版. 北京：高等教育出版社

董增福. 2003. 矩阵分析教程. 哈尔滨：哈尔滨工业大学出版社

方保镕，周继东，李医民. 2004. 矩阵论. 北京：清华大学出版社

黄廷祝，等. 2003. 矩阵理论. 北京：高等教育出版社

黄有度，等. 2004. 矩阵论及其应用. 合肥：中国科学技术大学出版社

罗家洪. 2000. 矩阵分析引论. 广州：华南理工大学出版社

卜长江，等. 2003. 矩阵论. 哈尔滨. 哈尔滨工业大学出版社

苏育才，等. 2006. 矩阵理论. 北京：科学出版社

徐树方. 1995. 矩阵计算的理论与方法. 北京：北京大学出版社

徐仲，等. 2001. 矩阵论简明教程. 北京：科学出版社

于寅. 1995. 高等工程数学. 武汉：华中理工大学出版社

曾祥金，等. 2007. 矩阵分析及其应用. 武汉：武汉大学出版社

Frankin J N. 1968. Matrix Theory//Englewood Cliffs. New York：Prentice-Hall

Jacob B. 1990. Linear Algebra. New York：L W. H. Freeman and Company

Lancaster P，Tismenetsky M. 1985. The Theory of Matrices Second Edition with Application// Orlando. Florida：Academic Press

[General Information]
书名=矩阵分析简明教程
作者=曾祥金，张亮主编
页数=221
SS号=12682986
DX号=
出版日期=2010.08
出版社=科学出版社

封面
书名
版权
前言
目录
第1章　线性空间与线性变换
1.1　线性空间的基本概念
1.2　子空间与维数定理
1.3　线性空间的同构
1.4　线性变换及其矩阵表示
习题1
第2章　内积空间
2.1　内积与欧氏空间
2.2　欧氏空间的正交基
2.3　欧氏空间的同构
2.4　正交补
2.5　正交变换
2.6　酉空间（复内积空间）简介
2.7　正规变换与正规矩阵
习题2
第3章　矩阵的标准形
3.1　Jordan标准形
3.2　λ-矩阵及其Smith标准形
3.3　Cayley-Hamilton定理与矩阵的最小多项式
习题3
第4章　矩阵分解
4.1　矩阵的LU分解
4.2　矩阵的QR分解
4.3　矩阵的满秩分解
4.4　矩阵的奇异值分解
4.5　广义逆矩阵
习题4
第5章　范数理论及其应用
5.1　向量范数
5.2　矩阵范数
5.3　范数的应用
习题5
第6章　矩阵分析及其应用
6.1　矩阵序列与矩阵级数
6.2　矩阵函数及其计算

6.3　矩阵的微分与积分
6.4　矩阵函数的应用
习题6
第7章　矩阵特征值的界非负矩阵
7.1　Ger？gorin定理
7.2　特征值估计的基本不等式
7.3　Courant-Fischer定理和Hermite矩阵的特征值
7.4　正矩阵
7.5　非负矩阵
7.6　随机矩阵
7.7　M矩阵
习题7
习题答案与提示
参考文献